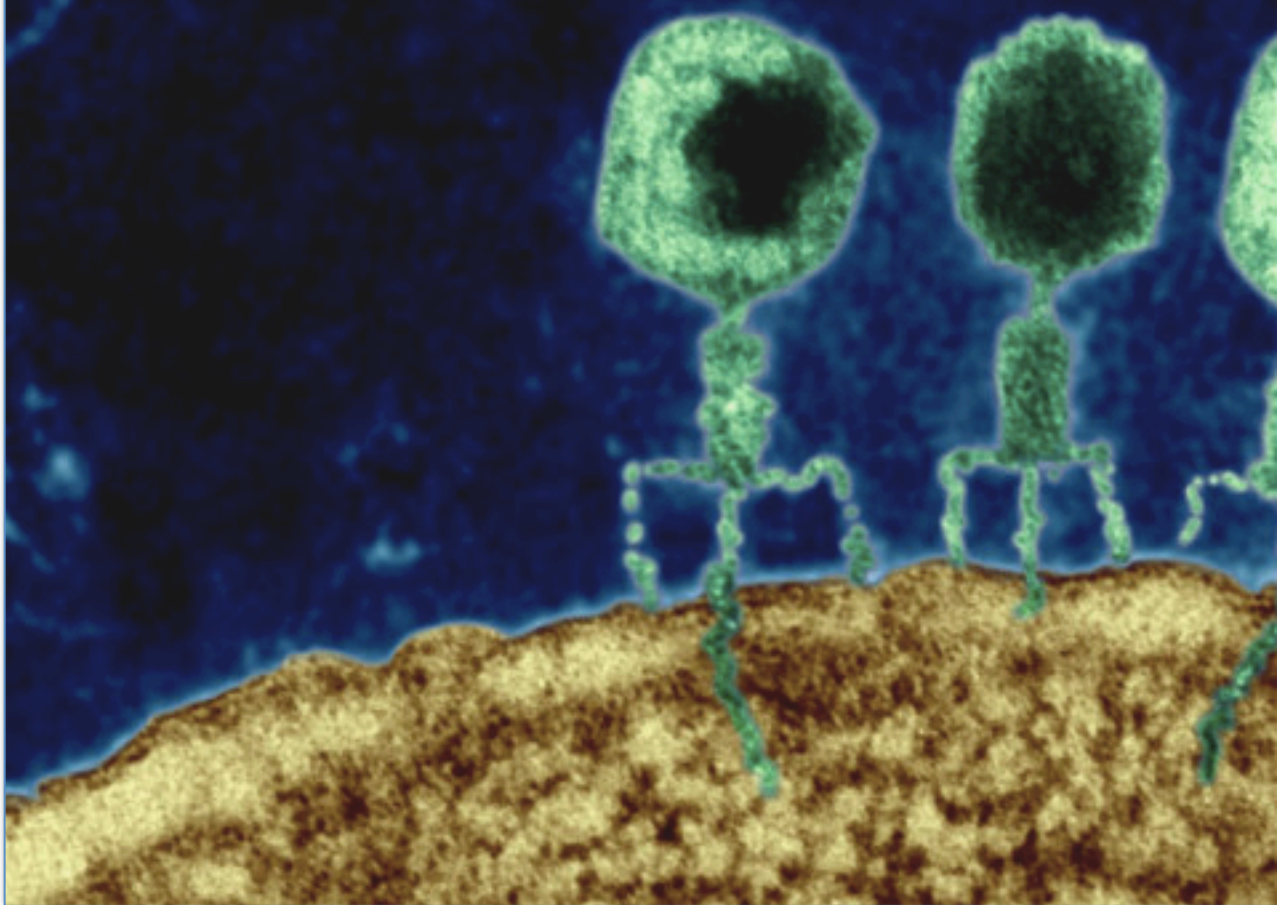
Acceleration of cell factories engineering using CRISPR-based technologies

Ronda, Carlotta

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Acceleration of cell factories engineering using CRISPR-based technologies



Acceleration of cell factories engineering using CRISPR-based technologies

ISBN:

Novo Nordisk Center For Biosustainibility
Department of System Biology
Technical University of Denmark

Who are *you? – Asked the Caterpillar -*
I hardly know, Sir I know ***who I was*** *but I think I must have changed – Alice -*

Lewis Carroll, Alice in Wonderland

Preface

This thesis is written as a partial fulfillment of the requirements to obtain a Ph.D. degree at the Technical University of Denmark. The work was carried out from June 2012 to May 2015 in the bacterial section at the Novo Nordisk Foundation Center For Biosustainibility, Department of DTU Biosustain, Technical University of Denmark under the supervision of Alex Toftgaard Nielsen (supervisor) and Søren Molin (head of section). The work was funded by Ph.D. stipend from Technical University of Denmark

---

Carlotta Ronda

Hørsholm, August 2015

Abstract

The constant demand of oil-derived products in the market has pushed science to develop alternative ways to cope with this demand. Therefore the development of efficient cell factories as sustainable alternative is an expanding trend. These are envisioned as future workhorse manufacturers of pharmaceuticals, biofuels and biomaterials. The focus of this thesis is to develop new genome engineering methods to relieve one of the major bottlenecks in metabolic engineering, the strain design and optimization. The aim is to generate an engineering tool-box applicable to different model organisms, which can potentially be standardized in an automatable platform and, in the future be integrated with metabolic modeling tools. In particularly it describes the technologies developed in the three widely used organisms: *E. coli*, *S. cerevisiae* and CHO mammalian cells using the recent breakthrough CRISPR/ Cas9 system. These include CRMAGE, a MAGE improved recombineering platform using CRISPR negative selection, CrEdit, a system for multi-loci marker-free simultaneous gene and pathway integrations and CRISPy a platform to accelerate genome editing in CHO cells.

Danske resumé

Det constant forøgede forbrug af fossile ressourcer til fået videnskaben til at fokusere på at udvikle alternative metoder til at fremstille de kemikalier, der er brug for i vores samfund. Fremstilling af biokemikalier ved hjælp af biologiske cellefabrikker er derfor et hastigt voksende forskningsområde. Disse cellefabrikker har mulighden for at producere fremtidens farmaceutiske stoffer, biomaterialer og brændstoffer fra genanvendelige ressourcer. Fokus i denne afhandling har været at udvikle nye metoder til at modificere genomer, sammen med design og optimering af produktionsorganismer. Målet har været at udvikle en sæt redskaber til modificering af genomer. Disse metoder har mulighed for at blive standardiseret i en automatisk platform, og som i fremtiden vil kunne blive fuldt integreret i de redskaber, som man bruger til metabolic engineering. I særdeleshed har fokus været på at udvikle teknologier til tre af de mest anvendte organismer, E. coli, S. cerevisiae samt de mammale CHO celler, ved at udnytte gennembruddet indenfor CRISPR/Cas9. De udviklede metoder inkluderer CRMAGE, som forbedrer effektiviteten af MAGE recombineering ved hjælp af CRISPR-baseret negativ selektion. CrEdit er et system til effektivt at integrere flere kassetter på genomet i S. cerevisiae uden brug af selektionsmarkører, mens CRISPy er en platform til at accelerere modifikation af genomer i CHO celler.

Contents

Acknowledgements

These past three years have been an emotional rollercoaster. I was caught in a very competitive field without either knowing where to start from or how to face such challenge alone. It was tough starting a Ph.D. project in a field about to bloom, although not to my knowledge at the time, without having someone already skilled in the field whom I could seek advices for. Despite that I have been very lucky to be given the chance to choose what I was passionate about without restraining my curiosity. For this reason I would like to truly thank my two mentors Prof. Alex Toftgaard Nielsen and Prof. Søren Molin who first believed in me when in early 2012 I prosed them to engineer CRISPR as genome editing toolbox and to modulate gene expression. Little was known at the time and the project seemed risky for a newly graduate, nevertheless they let me pursue my ideas and intuitions.

Being at CFB has been incredibly stimulating experience and I had the opportunity to interact with amazing and talented people. Among all I want to particularly thank Helene Faustrup Kildegaard, whom I had the great pleasure to co-author a paper and filed a patent with, for the great scientific discussions and personal talks that allow me to get to know how great of a person she is. To Lasse Edrupt Petersen, a talent and witty humorous researcher, who has been one of the first people I started the CRISPR adventure with. To also Jerome Maury whom I had the chance to collaborate with closely and to appreciate him both professionally and as a person.

I would like to express my gratitude to all my collaborators on my several projects. This includes all the co-autothors and the people I have been involved in projects with at CFB. It has been a great pleasure to work with all of you.

I want to include here my officemates, JiangFeng Zhu , such an amazing and clever scientist beside a great and genuine person who never refuses to help and answer even to the stupidest questions; and Torbjørn Ølshøj Jensen who has been kind enough to bear all my ranting. I would like to thank also the people in Alex's group, Christian Bille Jendresen, Sheila Ingemann Jensen and Hemanshu Mundhada for all the help and the great scientific discussions.

I would like to express my deepest gratitude to my friends and co-workers who had to hear my complains and helped me to go through this Ph.D. So all the bactory Ph.D. fellows with special regards to my friends Zyzy (Isotta D'Arrigo), Mimi (Mafalda Cavaleiro), Gigi (Gheorghe Borja), Alexey Dudnik, Stefano Cardinale and Joao Cardoso.

I would like to thank also my friends Fabio Brandellero, Giovanna Vettorato, Federica Dall'Igna, Stefania Gasparin, Jeanne Fernandez, Federica Belmonte, Sara Sanchez, Isabel Valdés Marin, Roberto Ferro, Matteo Calaon, Cristian De Gobba, Guido Tosello, Fabio Caponetti, Sasha Sansonetti, Benji Martinez, for supporting me and teaching me that work is not all in life. To also Andrea Mazzucco who has helped me to keep the

smile when the hell of the last few months was about to come. He always had a nice thought that could make my day.

To my parents Giovanna and Roberto, and brother Carlo who have always been encouraging and been there when I needed.

A special thank goes to Martin Kamensky who has been there the whole time with the ups and downs, he cheered me up and always believed in me despite my doubts. He was the only one who could calm me down and put me back on track, he was the sledge that safely drove me throughout this intense period.

In conclusion I want to thank the person whom I really owe everything, Prof. Alex Toftgaard Nielsen who has been more than supervisor, he became a mentor and a friend. He has been always inspiring, enthusiastic and dedicated despite the challenge of establishing a big research group of his own with several people fighting for his attention. I truly enjoyed our brainstorming when his passion and dedication come out crystal clear. Thanks Alex, really!

Publications included in this thesis

Ronda C., Maury J., Jakociunas T., Abdessamad Jacobsen S., Germann S.M., Harrison S.J, Borodina I., Keasling J.D., Jensen M.K., Nielsen A.T. **CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae.** *Microb. Cell Fact.* **14(1),** 97 (2015).

Ronda C., L.E Pedersen, H.G Hensen, TB.Kallehauge, M.J. Betenbaugh, A.T Nielsen, H.F. Kildegaard. **Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool**. *Biotechnol. Bioeng.* **111,** 1604–1616 (2014).

Ronda C., L.E Pedersen, A.T Nielsen. **CRMAGE: CRISPR Optimized MAGE Recombineering combined using web-based oligos design.** *Scie.Rep.* Submitted

Patent included in this thesis

Ronda C., Pedersen L.E., Kildegaard H.F., Lee J.S. **"Multiplex editing system**". EPO Patent Application (WO2015052231)

Abbreviations

Cas9: CRISPR associated gene

Cas9^{D}: nuclease defective mutant of Cas9

Cascade: CRISPR-associated complex for antiviral defence

CHO: Chinese Hamster Ovary (cells)

Cre: Cyclization recombinase or Cause recombinase

CRISPR: Clustered Regularly Interspaced Short Palindromic

CRISPR array: array containing repeat-target-repeat sequences

CrRNA: CRISPR RNA

CrRNP: CRISPR ribonucleoprotein

CRISPR-TFs: CRISPR- based Transcription Factor (dCas9 fused to effector domains)

DSB: Double Strand Break

F-LCA: fluorescein-labeled LCA

FLP: Flippase

FRT: Flippase Recognition Target

gRNA: guide RNA (synthetic)

HDR: Homology-Directed Repair

IgG: Immunoglobulin G

Indel: insertion or deletion

Kl.URA3: Kluyveromyces lactis URA3 gene

LCA: *Lens Culinaris* Agglutinin

MGE: Mobile Genetic Elements

MMEJ: Microhomology-Mediated end Joining

NHEJ: Non-Homologous End-Joining

PAM: Protospacer Adjacent Motif

Protospacer: The target DNA sequence that is complementary to the spacer of the CRISPR RNA.

ssODN: Single Stranded Donor Oligonucleotide

TALENS: Transcription Activator-Like Effector Nucleases

trascrRNA: Trans-Activating CrRNA

ZNFs: Zinc-Finger nuclease

List of figures

Chapter 1: Introduction and Thesis Overview

1.1 Cell Factories And Metabolic Engineering Bottlenecks

One cubic mile of oil (CMO) corresponds very closely to the world's current total annual consumption of crude oil. The world's total annual energy consumption from all energy sources is currently 3.0 CMO. By the middle of this century the world will need between 6 and 9 CMO of energy per year to provide for its citizens[1]. We are facing the necessity of producing the everyday-life oil derived goods in a cheap, fast and eco-compatible manner. For this reason in late 90´s, early 2000 the idea of engineering cell as sustainable alternative to produced chemicals and fuels has begun to grow and to gain importance in the scientific community. In a market of oil derived chemicals worth about US$ 5.5 Trilion in 2015, according to the New Report by Global Industry Analysts Inc., the potential of sustainable biotechnology is unlimited. Cell factories are envisioned as future workhorse manufacturers of pharmaceuticals, biofuels and biomaterials. Indeed the sources are infinitely renewable (cells replicate themselves) and the substrates can be adapted on demand by modifying the cell metabolism or by reconverting them into profitable substrates in microbial communities.

Metabolic engineering is a multidisciplinary field where several disciplines such as genetic engineering, molecular biology, biochemistry, system biology, chemical engineering collaborate in order to convert cells into

biocatalysts for a cost-effective production of chemicals[2]. Whilst this field is rapidly improving, there are still several issues that need to be addressed for a more competitive bio-based industry; such as the enhancement of the substrate spectrum with the focus on the efficient utilization of renewable feed-stocks, broaden out the product spectrum and hosts, improve genome editing and genetic tools to control cell behavior and maximize production[3]. This discipline was first introduced in the early 90's by Bailey as subgroup of the engineering field concerning of "the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology"[4]. Currently one of the main bottlenecks to achieve an efficient, competitive and versatile cell factory is hindered by the limitations of conventional methods of strain design.

In this scenario synthetic and molecular biology have the potential to provide tools for engineering cell factories in a more efficient and controllable way. Thus their improvement in the context of a more systematic approach with the advent of OMICS and computational modeling approaches will allow to expand the arrays of obtainable products and available hosts as well as achieve titers and yields more competitive for the global market.

The focus of this thesis is to create a set of genetic engineering tools applicable as potentially high-through put, parts-standardized and automatable platform for virtually any organism, both prokaryotes and

eukaryotes. In particularly, we developed genome editing methods using CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) technology in order to accelerate the strain construction.

1.2 Thesis Outline

This thesis is structured in seven chapters. The current one (**Chapter 1**) aims to introduce the field where this thesis has been focused on and the broad scope. The second chapter (**Chapter 2**) introduces the CRISPR system and describes the technology derived from it. Chapter three, four and five (**Chapter 3; Chapter 4; Chapter 5**) are organism-based genetic engineering monographs. They focus on the model organisms, *E. coli, S. cerevisiae* and CHO cells, respectively and the state of the art genetic engineering tools followed by the description of the CRISPR-based methods that have been developed during this project, and their contribution to the field. Finally chapter six (**Chapter 6**) discusses the future prospective and provides a final constructive conclusion of this thesis. Chapter 7 (**Chapter 7**) collects all full lengths papers.

Chapter 2: CRISPR revolution

The advent of recombinant DNA technology in the late 70s opened a new era for biology. For the first time it became possible to manipulate DNA and reprogram the genetic code of potentially any organism. It enabled the direct study of DNA in the context of cell physiology by modulating or editing the function of its sequences in their endogenous context. It was possible to elucidate the functional organization of the genome at the system level and the casual genetic variations. The recent advance of genome editing and engineering techniques due the advent of CRISPR/Cas technology integrated with next generation sequencing and gene synthesis, and the merging of large datasets with modeling and novel bioinformatics tools, has opened new prospective for the design of industrially relevant strains for production of novel compounds.

2.1 CRISPR history and phylogeny

CRISPR is a bacterial immune mechanism used by many prokaryotes to protect themselves from foreign nucleic acids like viruses or plasmids. Most of the Archea (about 84%) and bacteria (about 48%) bear this system in their genome[5]. It appeared in literature for the first time in 1987 described as "five highly homologous sequences of 29 nucleotides arranged as direct repeats with 32 nucleotides as spacing"[6], but it was only in 2002 that the name

CRISPR (Clustered regularly interspaced short palindromic repeats) was coined for the first time[7]. Initially it was inferred by computational studies that this system might be a possible repair system due to the high homology of the genes associated with it to helicases and nucleases[8], however later studies discovered that CRISPR spacers were homologous to fragments of mobile genetic elements (MGE) and thus they hypothesized for the first time that CRISPR system could instead be an adaptive immunity system[9–11]. However, it was not until in 2007 that Barrangou *et al.*, described its function as system of acquired resistance to bacteriophages via integration of viral DNA fragments in the lactic acid bacterium *Streptococcus thermophilus*[12].

CRISRP/Cas systems are extremely heterogeneous in terms of the repeats[13] and *cas* gene sequences beside the architecture of the *cas* genes within the operons[7,14,15]. This is probably due to the rapid evolution they have undergone because of the strong selective pressure imposed by the foreign MGEs. On the basis of this phylogenetic diversity they, are grouped into three major CRISPR/Cas systems and several different subtypes[16]. These three main systems are defined on the basis of the three unique signature effector genes: Cas3 nuclease-helicase for type I, Cas9 nuclease for the type II and for the type III Cas10 that so far has an unknown function. The type II is phylogenetically and structurally distinct from Type I and III, which seem more related[17]. Indeed while Type I and type III CRISPR–Cas systems are

spread in various combinations among phylogenetically diverse bacteria and archaea, the type II systems (sometimes in combination with other CRISPR–Cas types) are restricted to only few species of bacteria. Different subtypes of CRISPR/Cas system are classified by the way the crRNA and Cas proteins form the crRNP complex (CRISPR ribonucleoprotein), which is the machinery required to target and cleave invading nucleic acid[11]. The **cascade** (CRISPR-associated complex for antiviral defense) is the complex formed by the subtypes of type I (A to F) whereas in the type II (type II-A, type II-B and type II-C systems) the complexes are defined as Cas9. Type I and type III CRISPR/Cas systems are found in various combinations among phylogenetically diverse bacteria and archaea, whereas the type II systems, often in combination with other CRISPR/Cas types, have been found only in some bacterial species[14,18]. It has been speculated that the phylogenetic heterogeneity of the CRISPR/Cas system is the result of horizontal gene transfer[15], consistently with the fact that it has been found also in several viral genomes and plasmids[19–23].

2.2 CRISPR mechanism of action

The CRISPR/Cas surveillance system consists of three main stages: **adaptation**, **expression** and **interference**. Based on their function during the adaptive immune system, all Cas proteins can be grouped into four categories: (i) nucleases and/or recombinases, which are involved in the adaptation process; (ii) ribonucleases that process crRNA guides and (iii) the proteins that assemble with the RNA guides to form the crRNP complexes, which characterize the expression stage; and other nucleases (iiii) that are recruited to the complete crRNP complex to degrade the DNA or RNA in the final stage, in the interference[24] (Fig 1).

I. Adaptation: the adaptation represents the stage where the bacterium construct the "memory" of the virus or plasmids invaders by incorporating 30 nucleotides of their DNA at the leader side of the CRISPR locus[12,25,26]. The spacer acquisition takes place at the leader end and they are arranged as a chronological record of previously encountered invader nucleic acids. Therefore the closest are the spacers to the leader the most recent is their acquisition[12,27,28] (Fig. 1). The selection of the spacer is directed upon detection of a specific short nucleotide sequence known as PAM (Protospacer Adjacent Motif)[29–31] followed by processing of the DNA substrates into spacer precursors of a defined size[32]. Each new acquisition has to be accompanied by the duplication of the leader end repeat in order to create a new spacer-repeat unit[33] and the new spacer is integrated according

to the PAM orientation[32,34]. Occasionally a spacer derived form chromosomal DNA can be integrated but they are typically associated with toxicity[35,36] unless the PAM is mutated or Cas protein are inactive[33,35]. This proves that CRISPR/Cas systems can distinguish invading DNA from the host chromosome. Some of the mechanisms to avoid autoimmunity have been partially elucidated but most of them remain still unclear. It seems that the host restriction and methylation system might be involved[24].

II. crRNA processing: during this stage the crRNAs are produced from a long single transcript of the CRISPR array followed by a maturation process catalyzed by endoribonucleases that cleave the precursor long pre-crRNA into the final short crRNAs (Fig. 1). The enzymes involved in this process can be either a subunit of a larger complex (e.g. Cascade) or operate as single enzymes[37,38]. The primary processing of the pre-crRNA in type I (with the exception of I-C system[39]) and type III is catalyzed by Cas6 which typically leaves a 5' handle of 8 nucleotides a central spacer sequence and a longer 3' handle[37,40–42] that can either form a stem-loop secondary structure or undergo further processing, of which the mechanism of action and enzymes involved remain still unknown[38,43,44] (Fig. 1). The type II system relies upon a completely different mechanism based on Cas9 endonuclease and the indigenous RNAase III. This will be further discussed in the following paragraphed (2.3 Type II system).

III. Interference: during the interference the invading MGE (either double-stranded DNA or single-stranded RNA) is recognized and degraded by the surveillance crRNP complex (Fig. 1). This is a stepwise process that involves first the recognition of the protospacer sequence by base pairing between the 7–8 nucleotide seed region of the spacer and the complementary crRNA accompanied by the discrimination of self versus non-self DNA[45–49], followed by complete base pairing between spacer and protospacer that generates a strand displacement[46,48] and an R-loop structure. This seems to trigger a conformational change in the in the crRNP surveillance complex which leads to the intrinsic nuclease activity of type II and type III-B systems or promotes the recruitment of a trans-acting nuclease in the type I and type III A[40,50–52].

In order to not be deleterious and trigger an autoimmune response, the CRISPR system has to be able to distinguish between "self" from "non-self". *In silico* studies have elucidated that CRISPR surveillance system type I and type II rely on a specific motif of generally 2-5 nucleotides to discriminate the crRNA spacer on the host genome from the spacer precursor (proto-spacers) sequence of the invading DNA, which is the sequence complementary to the crRNA[31]. This particular motif is recognized as the Protospacer Adjacent Motif (PAM)[29,30] and it is located at the 5' end of the protospacer sense strand in the type I system, whereas it is at the 3' end in the type II system[29–31] (Fig. 1). The PAM has been demonstrated to

play a key role in the adaptation of type I[49,53,54] and type II systems[32,55–57], but it does not seem to be required in the type III system. Indeed while the first two systems have been shown to have a "non-self activation" strategy upon recognition of the PAM by the crRNP complex that triggers the interference by either recruiting a nuclease (like Cas3 in type I systems) or inducing the intrinsic crRNP nuclease activity (like in the case of Cas9 in type II systems)[48,58,59]. In the type III system this strategy is substituted by a PAM-independent mechanism of "self inactivation". This response is mediated by the interaction between the 5'-handle of the crRNA bound to the crRNP complex, and the repeat sequence of the host CRISPR array. The base paring of the repeat in the CRISPR array with the crRNA assembled to the crRNP complex activates the "self DNA" recognition thus preventing the interference most likely by blocking the recruitment of the endonuclease[60].

2.3 Type II CRISPR/Cas systems

While the type I and type III systems share some overarching characteristics, like specific nuclease-mediated maturation of the crRNA and a multi-subunit complex (e.g. cascade) that is recruited and guided by the mature crRNA to recognize (by sequence complementarity) and cleave the foreign nucleic acid, the type II system differs considerably in the mechanism of crRNA processing and interference. Indeed these two processes rely on three main components (i) a type II specific RNA, the trans-activating crRNA (tracrRNA), (ii) the crRNA, (iii) and a single nuclease Cas9 (formerly

Csn1). Cas9 is a large bi-lobe protein that contains two endonuclease domains (an HNH close to the N-terminal and an RuvC-like domain in the middle of the protein) and it is considered the major player in both the crRNA processing and the intereference[12,25].

II. crRNA processing: during pre-crRNA processing the tracrRNA (that contains 25 nucleotides complementary to the repeat sequences in pre-crRNA) forms a RNA duplex with the pre-crRNA repeats, together they assemble in the Cas9-RNP complex that promotes the pre-crRNA processing (probably by binding and positioning the RNA molecule) mediated by the double-stranded (ds) RNA-specific ribonuclease RNase III[18,56,61]. After the first processing the tracrRNA:crRNA hybrid remains docked to Cas9[55,56] and undergoes to further maturation at the 5'-end, resulting in a final spacer length of 20 nucleotide[61].

III. Interference: the hybrid tracrRNA:crRNA with the processed spacer length, in a complex with cas9, forming the Cas9-RNP is the only component required for the interference[25,55,56,61]. The PAM-dependent (GG dinucleotide-containing) target recognition mediated by the base pairing between the spacer and the foreign DNA, triggers a conformational change that activates the nuclease domains responsible for the interference. Recent studies have unraveled the Cas9 structure (from type II-A and C) both alone and in complex with the sgRNA (a synthetic RNA result of a fusion between tracrRNA and crRNA[56])[62,63] and has made it possible to elucidate the

mechanism of interference in more detail. Cas9 contains two highly conserved lobes: an α-helical recognition lobe, which is primarily involved in coordinating the guide RNA, and the nuclease lobe, which is responsible for PAM recognition and subsequent cleavage of the target. When the sgRNA is loaded to form the Cas9-RNP complex, it causes a conformational change in the protein making it ready to bind the target. Subsequently Cas9-RNP complex scans the DNA seeking for the PAM motif[58] located close to the 3' handle of the crRNA spacer region. After the PAM recognition and the subsequent target matching at 12 nucleotide seed sequence, the RuvC-like domain displaces the target DNA strand[56,62] and there is a formation of a R-loop structure that triggers the intrinsic nuclease activity in both HNH and RuvC domains of the nuclease lobe [62,63]. The HNH domain cleaves the DNA strand that interacts with the crRNA, while the RuvC-like domain cuts the displaced DNA strand. This results in a blunt double-stranded cut at a specific site, which is typically 3 nucleotides from the 3′ end of the protospacer[25,64]. Mutations in the PAM and seed region prevent the interference[65].

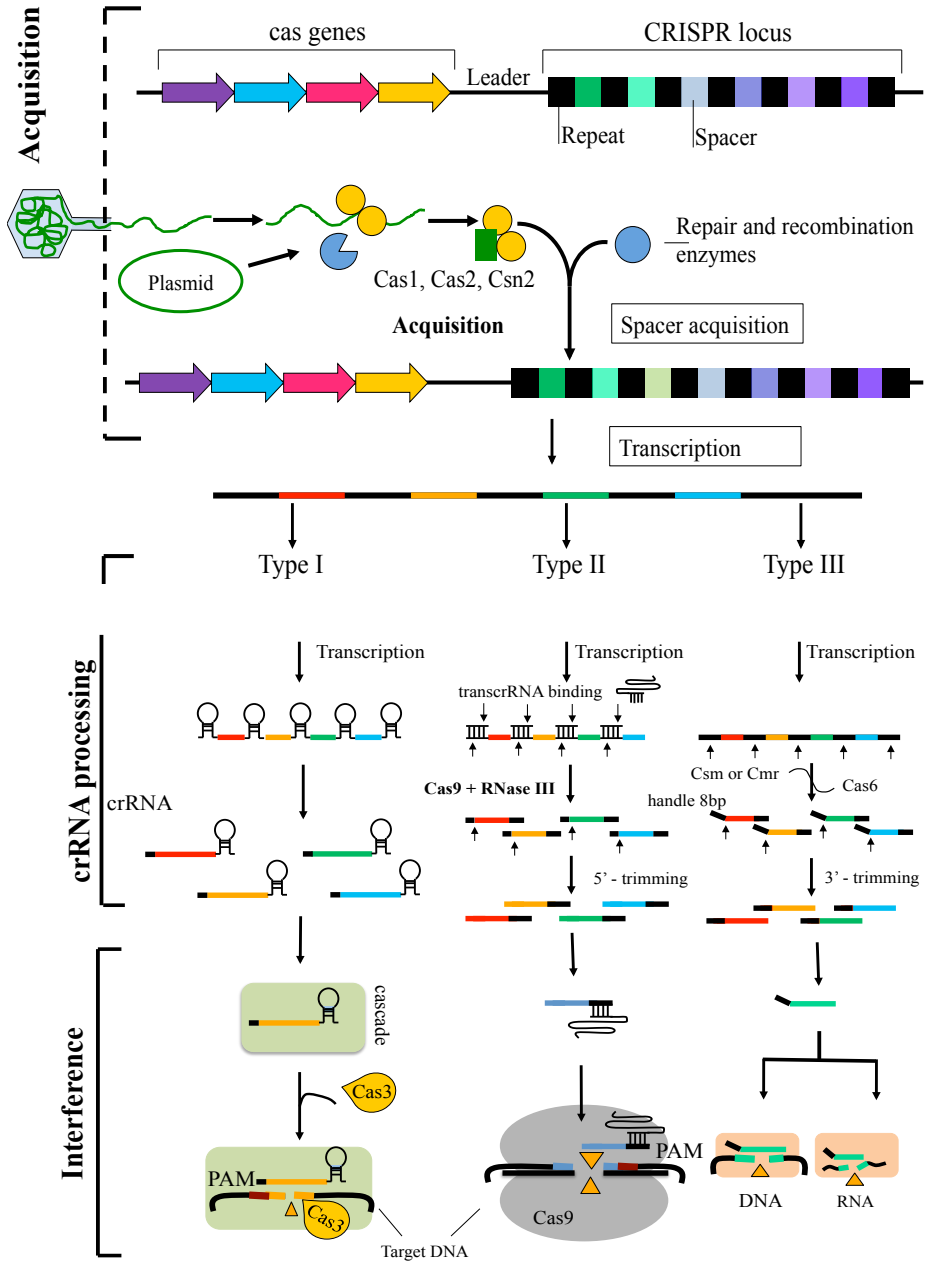


Figure 1: CRISPR mechanism of action overview : Here are graphically represented the three stages of CRISPR mechanism of action; (i) acquisition, (ii) crRNA processing and (iii) interference in the three major types.

2.4 CRISPR Craze: from immunity system to genome engineering tool

The remarkable ability of sequence-specific targeting and cleavage has made CRISPR/Cas an attractive system for genetic engineering. The understanding of the mechanism of action of the interference complex[66] and the mechanistic insights[58,61–63,67] has allowed scientists to reprogram CRISPR components, thereby opening a new revolutionary era of genetic engineering[68]. The Cas9 complex has been engineered as the "perfect tool" for genome editing by using synthetic sgRNA[56] (which is a fusion between trascrRNA:crRNA) or reprogramming CRISPR arrays[65] it has been possible to redirect this nuclease to induce DSB at almost any genomic locus. Mali *et al.*, and Cong *et al*., have independently demonstrated for the first time *in vivo* that it was possible to program the type II CRISPR/Cas9 nucleases (from *S. pyogenes*) using short synthetic gRNAs to induce precise cleavage at endogenous genomic loci in human and mouse cells[69,70], and shortly after Jiang *et al.*, showed it in bacteria using Cas9 from *S. pneumonie* and engineered CRISPR array[65]. CRISPR-based genome editing technology relies on the cellular DNA repair system that is triggered upon DSB introduction by the targeted nucleases. There are two different pathways in mammalian cells, non-homologous end-joining (NHEJ) or homology-directed (HDR) and each one of them gives different result when triggered by Cas9-generated DSB (Fig. 2A). The NHEJ is an error-prone repair system that involves direct ligation of the broken ends and can create indels

that result in frame-shift and consequent loss of function. For this reason, it is use for generating targeted gene knock outs (Fig. 2A). The HDR pathway instead uses homologous DNA sequences as templates for repair, and, by supplying an exogenous repair template, it can be exploited for precise genome editing or to insert exogenous DNA (Fig. 2A). The ability of Cas9 DSB-mediated highly efficient targeting of multiple loci at the same time in a cheap and easy way, especially for mammalian cell, has revolutionized the genome editing field and it has been applied successfully to basically any organism or type of cells from human cells [69–73], mammalia cell lines[74–77], mice[78–81], *Drosophila*[82–84], Zebrafish[85–90], *Xenopus*[91], *C. elegans*[92], silkworm[93], plants[72,94–96] and non- human primates[97] to bakery yeast[98–104] and bacteria[65,105–109]. Despite the great efficiency and possibility of multiplexing, concern has been raised regarding its specificity mainly due to its potential therapeutic implications. Therefore several studies have been focused on investigating further this aspect, part of the studies actually found some off-target effects but the results in different studies are incongruous (data ranges from 83% or 60% to less than 1%) so the whole-genome target specificity and the influence of chromatin structure on Cas9 acivity remain still controversial [58,110–118]. One strategy that has been applied to limit the off target effect was to use a mutated version of Cas9 that instead of producing DSB, cuts only one strand ("nicks" the DNA) (Fig. 2B). This should promote the HDR or base excision repair system (BER) over the NHEJ and

thus ensure high-fidelity repair and minimizing off target effect[70,119–121]. The Cas9 nickase is derived either from the mutation of the aspartate residue (D10A) in the catalytic site of the RuvC domain[70] or the histine (H840A) in HNH domain, depending on which strand is targeted[56] (Fig. 2B). Moreover, scientists inspired by previous genome editing tools such as TALENS and Zinc Finger nuclease, have fused a catalytically inactive Cas9 (mutated at both sites, D10A and H840A) to a Fok1 nuclease domain in order to further narrow its specificity (Fig.2D). Indeed it relies on the specificity of two different targets[122] instead of one since it is necessary that two Cas9 proteins are bound in close proximity to enable the interaction between the Fok1 domains and thus activate the nuclease activity (Fig.2D). In this way the probability of having a-specific binding and cleavage is sensitively reduced[122].

Using CRISPR/ Cas9, it has been possible to perform *in vivo* genome-wide functional genomic studies in eukaryotes by using full genomic libraries of gRNAs in different mammalian cell lines (human, mice, ESC) to identify genes associated with specific phenotypes[73,123–125] or to map tumor growth and metastasis[126]. Even though targeted gene knock out was already possible with other DNA binding nuclease (Zinc Finger and TALENs) [127,128] and RNA interference (RNAi) knockdown[129] was already employed for rapid, inexpensive and high-throughput method for genome-wide studies, CRISPR/Cas9 based approach has outperformed them qualitatively. Indeed

it enables more efficient multiplexing knockouts (which is basically impossible to perform using other techniques) and to generate more reliable and consistent data than RNA interference (RNAi) which varies between experiments and laboratories and has more unpredictable off-target effects[130], while providing only temporary inhibition of gene function[129].

The possibility of creating a DSB break virtually at any locus in genome has been exploited in bacteria as a negative selection system (Fig. 2A). For example Gomaa *et al.* have used the nuclease activity of Cas9 to program DSB-induce killing towards a specific sequence in the genome[131]. In this way they were able to selectively remove a population from a mixed culture. By using the same principle engineering techniques like λ red in *E. coli* has been improved and optimized (see **Chapter 3**). Moreover, Cas9 programmable toxicity has been used to develop sequence specific antimicrobials, whose spectrum of activity is chosen directly by the sequence of the gRNA-guide[132,133] or indirectly using lytic phages, by targeting the genes that confer resistance and thus enriching the antibiotic-sensitive population[134].

The remarkable ability to redirect a DNA binding protein just using a short RNA guide has made CRISPR/Cas9 technology appealing as a scaffold for various other applications. For example catalytically inactive Cas9, dead Cas9 (dCas9 or cas9D[9]), was generated by inserting two point mutations in the HNH (H840A) and ruvC (D10A) domains and used for targeted

silencing of transcription in bacteria and eukaryote cells[135–138] (Fig.2C). This system known as CRISPR interference (CRISPRi) is based on dCas9 co-expressed with a guide RNA to generate a DNA recognition complex that can specifically interfere with transcriptional elongation (Fig. 2C). When targeted to a promoter or ORF of a gene of interest, dCas9 can block progression of RNA polymerase and hence silence expression of the targeted gene. This technology has been proven to match perfectly the concept of synthetic cells, where the metabolic network can be redesigned using Boolean language and terminology known from electric components. As proof of concept Nielsen *et al.* have recently used CRISPRi system to construct genetic logic states in *E. coli.* Using a layered Boolean-logic of series of NOT and NOR gates, they were able to invert transcriptional inputs with >50-fold dynamic range thus demonstrating that CRISPRi can also be configured into transcriptional cascades that propagate signals as sgRNAs[139] (Fig.2C).

Due to its programmable DNA binding proprieties, dCas9 is amenable to creating fusions with effector domains (KRAB repressor and VP64 activator domains) and RNA modules to turn transcription on and off in a dynamic and quantitative manner, enabling the investigation of gene functions as well as rewiring regulatory networks[138,140,141] (Fig. 2E). CRISPR-TFs (CRISPR-based Transcription Factors) enable the construction of complex large-scale synthetic circuits and transcriptional cascades in order to redirect natural

regulatory networks and to create multi-component genetic circuits whose feedback loops, interconnections, and behaviors could be revwired[142]. In this prospective Zalatan *et al.*[143] have recently demonstrated the potential of CRISPR-TFs technology as scaffold to engineer complex synthetic transcriptional programs by attaching multiple protein-binding aptamers (MCP, PCP and com) to the sgRNAs linked to different aptamer structures (MS2, PP7 and com), for genomic localization of multi-protein complexes (CRISPR-TFs containing activator or repressor domains linked to MCP, PCP and com aptamers binding domain) functioning as genetic switches (Fig. 2F). Therefore, modified gRNA scaffolds can be used to generate synthetic multi-gene transcriptional programs in which is it possible to differentially switch on or off different genes at the same time (Fig. 2F). In conclusion, CRISPR based technologies have thrived in an efficient high-throughput screening system, especially for complex organism like eukaryotic cells, to interrogate genes functions and map pathway complexity by the perturbation of their expression. This system was successfully applied to screen for cancer drug resistance, tumor suppressors, genes related to growth defect, regulators of differentiation and genes involved in bacterial induced toxicity such as for anthrax, diphtheria and cholera [73,123–125,144,145].

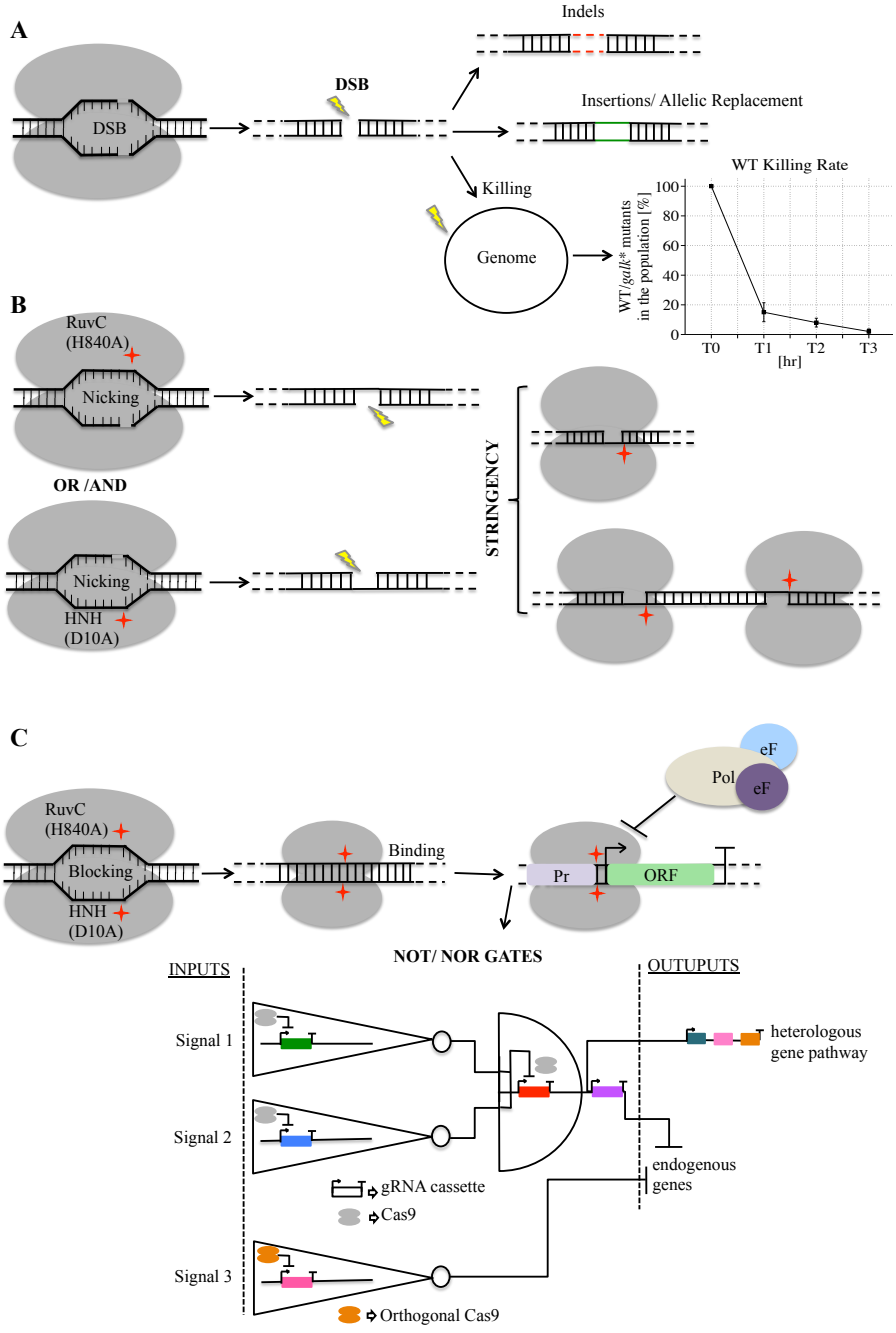
Especially CRISPRi and the derived CRISPR-TFs technologies can be utilized as a powerful tool for interrogating, perturbing, and engineering cellular systems. In order to expand the range of possibilities and create even

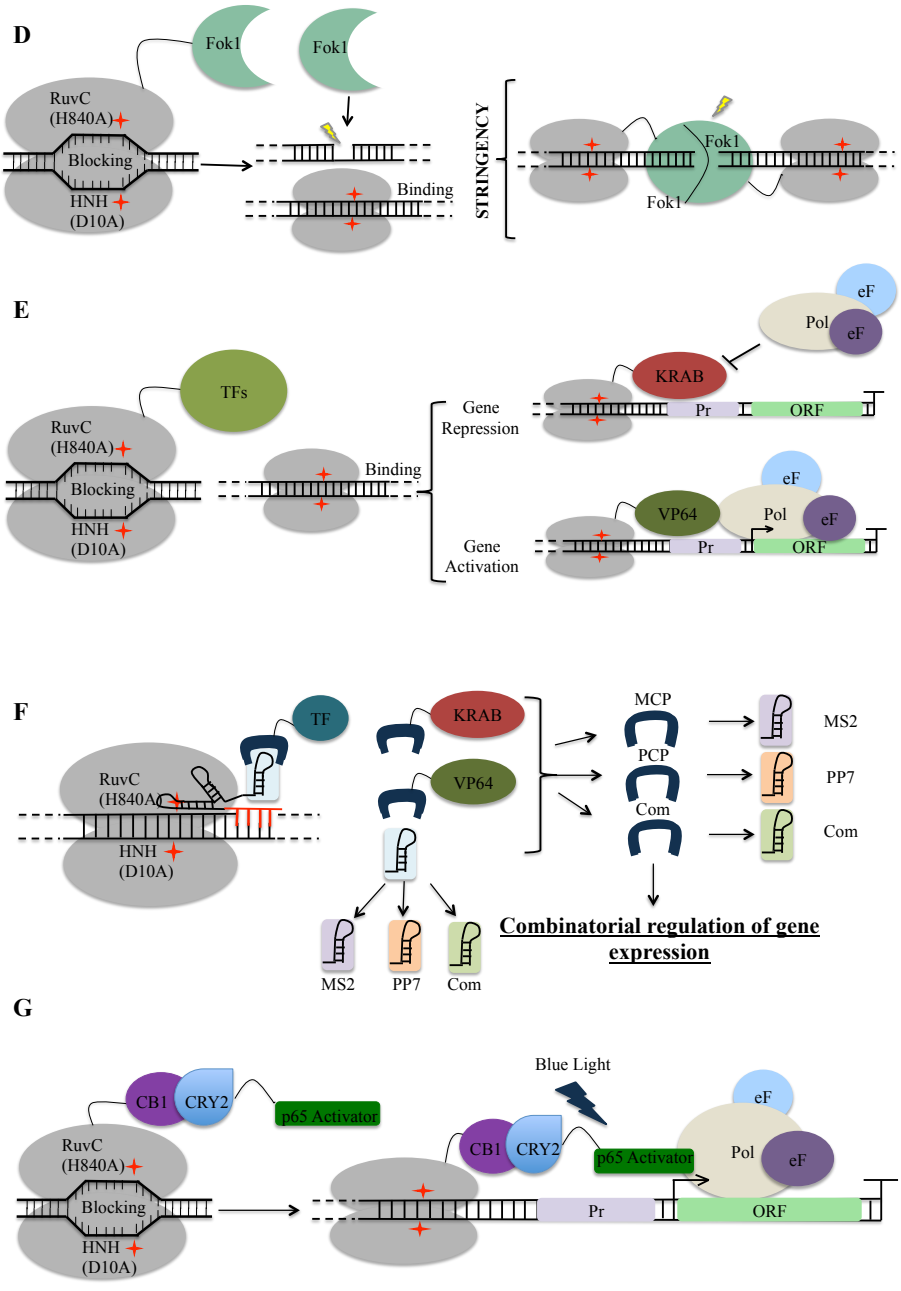
more architectural complex and articulated synthetic circuits, they could also be combined with orthogonal Cas9 [146] (Fig. 2C).

It has also been possible to modulate transcriptional activation by manipulating the chromatin architecture. Indeed Hilton *et al.*[147] have shown that by modifying the epigenome using a programmable dCas9-acetyltransferase fusion, it was possible to activate specific promoters via their enhancers (Fig. 2H).

Recently, also optogenetics has been integrated with CRISPR/Cas9 system to create a spatial and temporal transcriptional control[148] and genome editing[149] (Fig. 2G).

Besides the use of dCas9 as regulatory player in complex systems, it has also been employed as an imaging tool. Indeed its fusion to a fluorescent proteins[150–152] has created a novel *in situ* imaging platform that allows to better understand intracellular distance between loci on different chromosome and the spatiotemporal organization and dynamics of chromatin which plays a crucial role in regulating genome function (Fig. 2I).


A
Indels
DSB
DSB
Insertions/ Allelic Replacement
Killing
Genome
WT Killing Rate
WT/galk* mutants in the population [%]
100
80
60
40
20
0
T0
T1
T2
T3
[hr]
B
RuvC (H840A)
Nicking
OR /AND
Nicking
HNH (D10A)
STRINGENCY
C
RuvC (H840A)
Blocking
HNH (D10A)
Binding
Pol
eF
eF
Pr
ORF
NOT/ NOR GATES
INPUTS
OUTUPUTS
Signal 1
Signal 2
Signal 3
heterologous gene pathway
endogenous genes
gRNA cassette
Cas9
Orthogonal Cas9


D
Fok1
Fok1
RuvC
(H840A)
Blocking
HNH
(D10A)
Binding
STRINGENCY
Fok1
Fok1
E
TFs
RuvC
(H840A)
Blocking
HNH
(D10A)
Binding
Gene
Repression
KRAB
Pr
ORF
eF
Pol
eF
Gene
Activation
VP64
eF
Pol
eF
Pr
ORF
F
TF
RuvC
(H840A)
HNH
(D10A)
KRAB
VP64
MCP
PCP
Com
MS2
PP7
Com
MS2
PP7
Com
Combinatorial regulation of gene expression
G
CB1
CRY2
p65 Activator
RuvC
(H840A)
Blocking
HNH
(D10A)
Blue Light
CB1
CRY2
p65 Activator
eF
Pol
eF
Pr
ORF

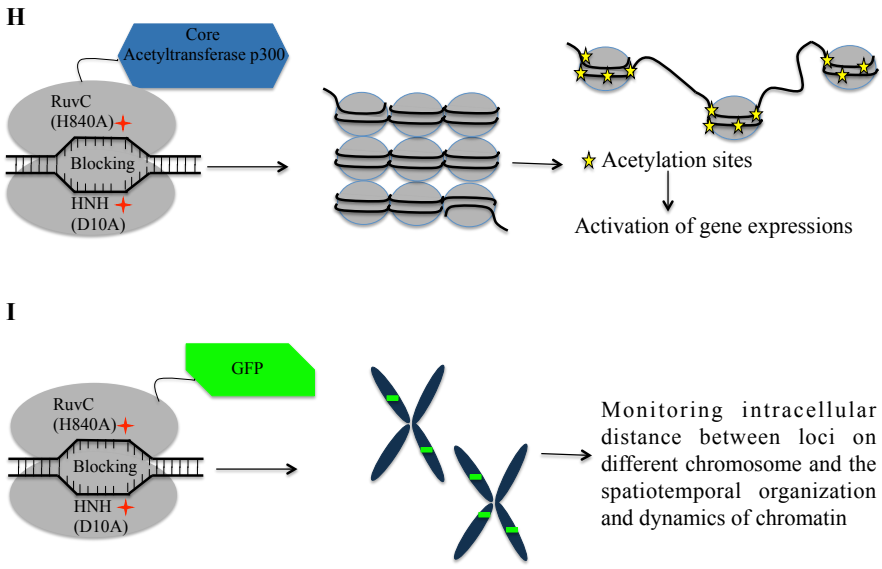


Figure 2: CRISPR/Cas9 applications overview. A: Cas9 nuclease applications. DSB mediated modifications (indels, Knock-in/out) and toxic activity used for specific negative selection. **B:** mutated Cas9 with only nickase activity to increase the target stringency. **C:** CRISPRi applications. The ability of dead Cas9 (double mutant) to bind any sequence in the genome with lack of nuclease activity is used to control gene expression. dCas9 is directed to the promoter region or the at the transcriptional start site to interfere with the polymerase and thus with the elongations of the transcript. It can be exploited for the construction of complex genetic circuits even using orthogonal Cas9. **D:** dCas9 can be associated with the nuclease domain Fok1 in order to increase the specificity. **E:** dCas9 can be fused with activator (PV64) or repressor (KREB) domains in order to control the gene expression. **F:** complex transcriptional circuits can be design by using dCas9, gRNA scaffold and aptamer binding proteins fused to activator or repressor domains. **G:** optogenetics can be integrated with CRISPR/Cas9 system to create a spatial and temporal transcriptional control and genome editing. **H:** it's possible to modulate transcriptional activation by manipulating the chromatin architecture using dCas9 fused to an acetyltransferase domain. **I:** dCas9 can be fused to GFP and directed to specific sites along the chromosomes. In this way dCas9 offers an *in situ* imaging platform that allows to better understand intracellular distance between loci and the spatiotemporal organization and dynamics of chromatin.

Even though the potential of CRISPR type II (Cas9) seems endless and even though it has already opened up complete new branches of genome engineering research, other CRISPR types or subtypes have shown to be amenable for engineering purpose. For example Qui *et al.*, have demonstrated the programmability of csy4 ribonuclease of CRISPR type III from *P. aeruginosa*[153] to create a synthetic RNA-processing platform for predictable gene expression by cleaving precursor mRNAs. Using this approach, it is possible to minimize the polar effects stemming from gene localization and thus achieve predictable regulation of multi gene operons[154]. This "synthetic splicing" has also been combined with CRISPR-Cas9 for gRNA multiplex processing[122], and for the construction of more complex circuit systems[142]. Furthermore the ability of Csy4 to bind RNA has been exploited to study specific RNA-binding proteins, and it has been suggested to improve the sensitivity of transcriptome studies by fishing multiple specific transcript using Csy4[155].

Also another type of CRISPR/Cas system has been engineered; indeed Luo *et al.*, have programmed the *E. coli* endogenous CRISPR/Cas system type I-E to modulate gene expression. By deleting the nuclease Cas3, which is the subunit of the cascade complex accustomed for the DNA degradation, it was possible to control gene repression and thus create an alternative, species-specific CRISPR/Cas regulator[156].

Chapter 3: *E. coli* improved recombineering

3.1 *E. coli* state of art engineering techniques

Over the decades *E. coli* has become one of the most widely used model organism for metabolic engineering. Thanks to the detailed knowledge about its metabolism it has been possible to reconstruct a fairly accurate metabolic network that has empowered constrain-based models able to simulate *E. coli* physiology by predicting the flux distribution and the possible products out come[157–159]. The genetic of this bacterium has been extensively studied and several methods to engineering it has been developed and optimized. *E. coli* is one of the most known model organism in metabolic engineering for production of a wide range of bio-base compounds[160]. The integrated use of systems biology, synthetic biology and evolutionary engineering has enabled an extensive development of genetic tools and protocols for efficient, fast and cheap manipulations in order to make *E. coli* suitable for industrial applications. Engineering methods have been evolved from simple allelic replacements and P1 trunsduction to the use of special enzymes as integrases, recombinases and special phage proteins for recombineering (Fig. 3). Moreover, repurposing of the retrohoming group II introns and the new breakthrough CRISPR technology have become a promising and successful methods to introduce modifications and rewire *E. coli,* and in general, bacterial genomes (Fig. 3).

Allelic replacement and P1 transduction: allelic replacement basic principle is the idea of creating genome modification by using the endogenous repair system to replace genomic sites with a new sequence by using long homology arms (Fig. 3A). This can cause gene disruption if the new sequence is incorporated within an ORF and usually it relies on a "suicide", non-replicative Ori plasmid carrier. The P1 transduction process is based on the inaccuracy of P1 phage packaging, which during the encapsulation of its DNA occasionally also incorporates pieces of its bacterial host. In this way when a P1 lysate is made, some of the particles also contain the host bacterial DNA, and when a second host is infected, then those pieces of bacterial DNA can be transferred into the new host through homology recombination[161] (Fig. 3A). This method can be useful only if the targeted alleles in the donor host genome are impaired or have antibiotic markers within them. Moreover the donor and the recipient host need to be closely phylogenetic related in order for the HR to occur.

Group II introns: mobile group II introns have been retargeted for chromosomal gene disruption and modification in a broad range of bacteria[162–164]. These bacterial retrotransposons work in a ribozyme manner to mediate a site-specific DNA integration ("retrohoming") through a retrotranscription step (Fig. 3B). Since most of the DNA target site is recognized by base pairing, the intron-insertion site can be programmed by simply modifying a specific motif in the intron RNA. The high efficiency

and easy programmability of the group II introns have enabled the construction of specific gene-targeting vectors, or "targetrons", for fast an easy manipulation that combined with Cre/Lox system, broaden the range of possible modifications[165,166]. Targetrons are widely used for genetic engineering of bacteria with an efficiency that vary from 1 to 80% depending on the site and species[167], and several efforts have been invested and still continue to adapt them for function in eukaryotes[168]. Moreover, the system is active in a wide variety of microbes, providing genetic manipulation of species that cannot be modified using other methods[165] .

Recombinase: recombinases are phage or yeast derived enzymes that can efficiently excise DNA between two target sites in direct orientation or an inversion of the sequence in between can occur if the target sequences are inverted as can co-integration of two separate plasmids that each carry a target site, albeit at lower efficiency[169]. There are three major recombinase enzyme: the λ phage recombinase using attP-attB sites target, the yeast derived Cre which recognize LoxP sites, and the yeast mitochondrial Sce-I which is specialized in scar-less excisions. In the context of genome engineering they are utilized to introduce large DNA constructs, to recycle the selective markers after recombineering in sequential allelic exchange, "Knock-out or knock-in of DNA[170–172] and for DNA shuffling[173] (Fig. 3C-D).

Recombineering: it is defined as genetic eng*ineering* using phage-derived *recom*bination proteins[174], and it has become probably the most widely used system for mutagenesis in *E. coli* because it is fast, simple and cheap since it relies basically on PCR products (for dsDNA) and DNA oligonucleotides (for ssDNA). The recombination system uses either the λ Red system or the RecET genes of the Rac prophage to dramatically improve the frequency of homologous recombination (HR) since the endogenous system, which relies on RecA/Rad5, works at very low efficiency and requires 4.5 kb to function properly. The high recombination efficiencies can be exploited in a variety of ways including the construction of chromosomal gene knock-out, knock-in, point mutations, genomic libraries, mutagenesis of bacterial artificial chromosomes and *in vivo* cloning[175–185]. One of the first methods developed that marked a turning point in recombineering technology it was the repurposing of rec ET proteins from the Rac prophage to manipulate *E. coli* genome and plasmid using sequential steps of homologous and site-specific modification. In their work Zhang *et al.*,[186] proposed for the first time a new concept of DNA engineering that moved forward from the classic restriction and ligation based methods. This system was based on Rec ET system and it was transferable between different *E. coli* strains. Despite the relatively low efficiencies, λ Red system has been the most employed and further developed recombineering method compared to Rec ET. It consists on three phage-encoded proteins, Exo (α), Beta (β), Gamma (γ) and it works

similarly to Rec ET system. Exo is a 5′→3′ exonuclease that degrades the dsDNA ends exposing a 3′ single-stranded DNA (ssDNA) tail to which the β protein can bind[176–178,187]. Subsequently the β protein-DNA form a complex with the target sequence through base pairing complementarity and promote recombineering by strand exchange or strand invasion[188–191], while the protein Gamma inhibits the degradation of dsDNA by binding the RecB protein of the RecBCD system[192,193] (Fig. 3D). RecE and RecT are functionally equivalents to Exo and Beta respectively[176]. λβ is a single-stranded DNA annealing protein[194] and it is the only component required for recombineering with ssDNA[183], while α and γ are necessary when dsDNA is involved. Phage-based recombineering is proficient with as little as 30-50 bp of homology flanking sequence to the target site when using PCR products containing selective markers or heterologous genes (typically an antibiotic resistance or metabolic gene) and 70 bp are considered the optimal length for ssDNA recombineering[177,178,183,195]. However, in order to select the recombinant clones, all these methods rely upon antibiotic marker selection which has to be removed in order to introduce further modifications. In order to recycle the antibiotic markers and allow sequential modification, Flp flippase (with FRT target) and Cre-LoxP mediated site specific recombination have been used for precise excision of selection markers[175,186,196–198]. Despite the great advantage of re-using the selective marker for the subsequent modifications, these site-specific recombination

systems leave at least a copy of the FRT site or the loxP site (scar) after excision of the markers, which limits the repeated use of this procedure since internal chromosomal rearrangements can occur. Therefore, new methods based on cam-sacB counter selection or Sce-I scar-less meganuclease have been developed in order to overcome the problem of genome instability associated to having multiple scars with the same sequence spread along the genome[179,184,195,199–203]. Notwithstanding the issues related to the recombinogenic scars, a simple PCR product with flanking homologous sequences to the target site or single stranded oligonucleotides, enable limited rewriting of any region of the genome.

Despite the good efficiency of the counter-selection, the engineering process becomes laborious and time consuming with multiple steps involved for the introduction of just a single modification. Additionally, the methods do not allow multiplexing since only few selective markers are available for combinatorial use and recombination with linear DNA in *E. coli* is limited because the transformed linear DNA is rapidly degraded by the bacterial RecBCD nuclease[204]. It has been demonstrated that it is possible to increase the number of recombinants by using mismatch repair mutants such as *mutS*[205] and by using phosphorothioate linkages to protect the lagging-targeting[206] strand in order to increase the half-life of the oligos or PCR products and thus the probability of their incorporation during the replication. The short ssDNA/dsDNA

oligonucleotide-mediated recombineering is best suited for inserting single genome mutations, and it turns challenging for targeted multiple gene insertions over a certain length[207]. Additionally, it becomes very difficult to screen for the desired mutations/indels if these do not display a clear phenotype.

CRISPR/Cas9: the type II CRISPR/Cas system from *Streptococcus pyogenes* or *pneomonie* has been recently proved to be a powerful tool for genome engineering in different Prokaryotes such as *E. coli* [65,106,107], *Actinomycetal spp.*[105], *Streptomyces spp*. [108,109], lactic acid bacteria[208]. The programmability of Cas9 has been shown to be a promising tool to manipulate genomes that are difficult to engineer, especially those where a recombineering system is absent or inefficient (Fig. 3E). In *E. coli*, the CRISPR/Cas9 system has been recently coupled to λ-RED oligo recombineering in order to improve its efficiency[65,106,107]. The power of CRISPR/Cas9 system is the ability to create a selective pressure during the recombineering procedure. Indeed the nuclease activity is retained towards the loci that have not incorporated the mutation thus constantly inducing double strand break that, in turn, cause cell death or a fitness defect (Fig. 3E).

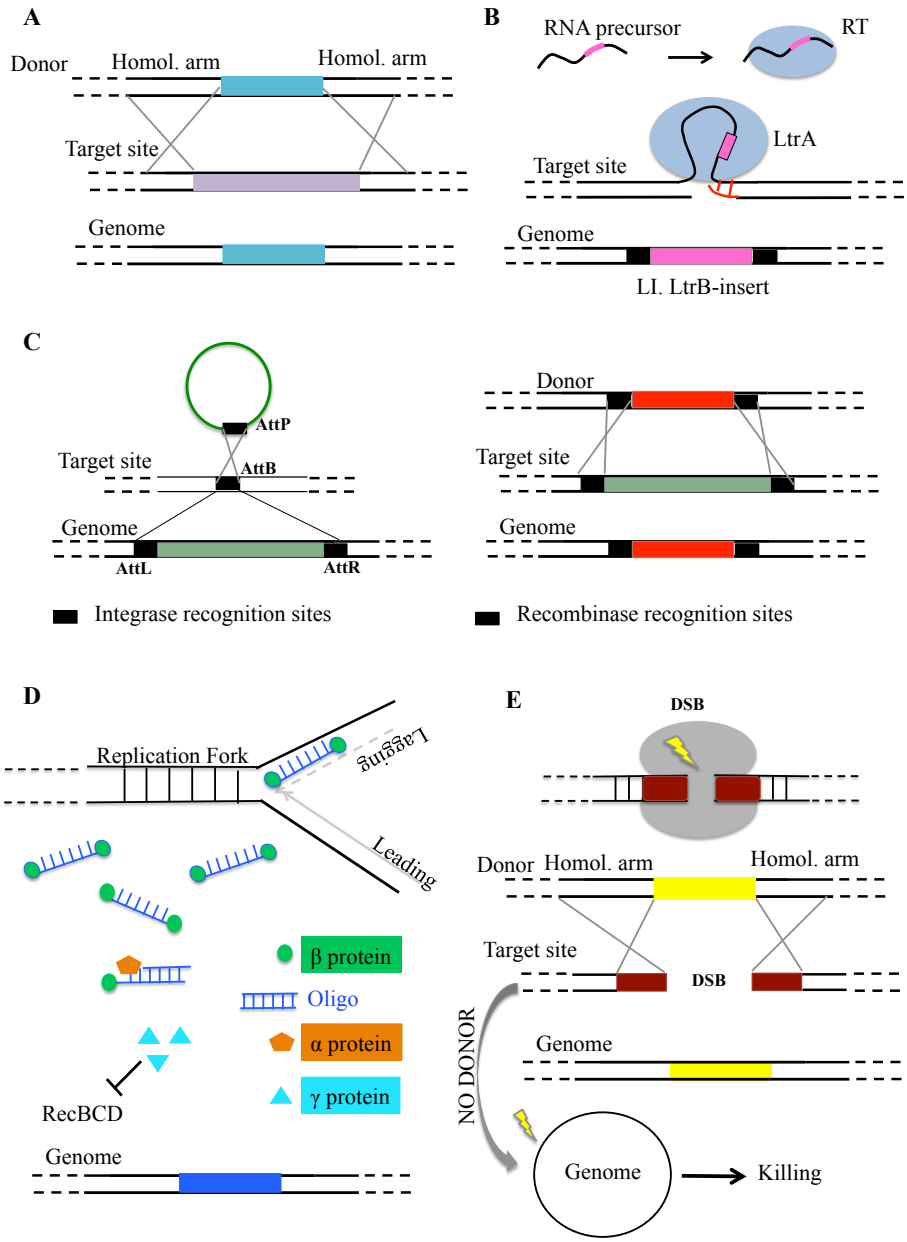


Figure 3: *E. coli* engineering methods overview. **A:** homologous recombination mediatetd modifications (can be knock-out/in or allelic replacement). P1 transduction relies on the same principle to insert the donor in the host genome. **B:** group II retrohoming can be programmed for site-specific genome insertion. **C:** recombinases and integrases allow DNA insertion or inversion between the two recognition sites. They

are often used for marker recycling after the modification has been inserted and selected for. **D**: λ red oligo mediated recombination. ssDNA or dsDNA oligos are used to modify the genome at a specific site by incorporating them at the lagging strand of replicating DNA. After the replication and cell segregations the mutation is inherited and fixed in the genome. **E**: Programmable endonucleases mediated modification. The system relies on the programmability of the endonuclease to induce DSB at specific genomic sites, which are then resolved with homologous recombination and insertion of the donor DNA. If the donor DNA is not present, constant introduction of DSBs can cause cell death. This can be used as programmable negative selection system.

3.2 Towards genome-scale engineering

In recent years the attention has been moved towards genome-scale approaches, especially thanks to the continuously decreasing in price and the high-throughput scale of DNA sequencing and DNA synthesis[209], recent breakthroughs in genomics and genome editing, new *in silico* systems for large data analysis and larger metabolic and signaling network models enable rational reconstruction of cellular behavior. Together with the growing field of synthetic biology, this promises a greater role for rational design by reverse engineering and synthetic cell circuits to reconstruct cell factories. Although the plethora of *in silico* metabolic and flux models promises reconstruction of virtual synthetic cells optimized for production, the metabolism and its intricate networks have not yet been fully elucidated, and we are still limited by our ability to construct the predicted genome. The absence of molecular tools for genomic sequence manipulations in a fast and cheap way is the major driving force that has led us to often rely on selective breeding and evolutionary optimization using ALE (Automated Laboratory

Evolution)[210] rather than rational genome design[211]. The restrictions that come with ALE are that evolution is biased by the first mutations that appear, which is not always the optimal path for pushing the cells towards the boundaries as would happen in cell factories. Thus, it does not sample all the possible solutions and most of the time it is difficult to find the selective pressure necessary for rearranging the flux towards production. Moreover, the objective function for ALE is growth coupled to production, which is not always optimal since a faster growing cell may be wasting energy on growth rather than production. In the best scenario when ALE would be able to evolve a strain with the desired phenotype, we are still facing the challenge of reverse-engineering the strain by inserting all the target mutations since there are no easy tools for identifying the causative mutants unless they are linked to a specific phenotype.

Recoding organisms at the genome-scale level requires methods that enable the manipulation of living cells at the whole-genome level. Such methods are increasingly becoming attainable through the advent of advanced technologies for manipulating, recombining and synthesizing DNA[212]. This has engendered to a serial of impressive genome-scale methods, which include genome merging[213], and recoding[214], whole-genome transplantation[215], refactoring phage genomes[216], creation of a bacterial cell controlled by a chemically synthesized genome[217], whole-genome synthesis [218], Tn5 whole-genome mutagenesis and oligos screening[170,219,220] and

removing transposable elements[221]. MAGE (Multiplex Automated Genome Engineering (MAGE)[222,223] was developed as a method that facilitates genome-scale engineering. Tractable Multiplex Recombineering uses barcoded primers derived from DNA microarrays to generate pools of dsDNA cassettes that can target different sites across the genome[224], while the CAGE (Conjugative Assembly Genome Engineering) technology permits genome-scale assembly of large portions of the genome via conjugation. This technology combined with MAGE, makes it possible to re-write the whole genome from single changes of genomic sequences with MAGE, to their assembly into a new chimeric genome[225]. Engineering of biological systems can be unpredictable and challenging since it often requires simultaneous genome-wide changes that are difficult to screen for. MAGE is the only method so far that allows multiplexing in one step by direct electroporation of SS-Oligos or PCR products. The efficiency of MAGE is highest for short genome modifications, where around 6 to 20 % can be achieved after single or multiple cycles respectively[183,222], while larger modifications occur with significantly lower frequency (< 1%)[206,222]. Several methods have been developed to improve the efficiency as the co-operative oligonucleotide co-selection[226] or the use of using phosphorothioate linkages to protect the lagging-targeting strand in order to increase the half-life of the oligos or PCR products and thus the probability of their incorporation during the replication[206]. However, the efficiencies

still remain relatively low, and when the desired mutations do not result in a clear phenotypic, then it is necessary to implement cumbersome screening such as by PCR[183]. To address this we have developed an approach that integrates MAGE recombineering method with CRISPR technology, CRMAGE, as described below.

3.3 Impact of CRISPR on *E. coli* genome editing: CRMAGE

The recent tendency in genome-scale engineering is to explore more complex network and cellular behavior at a whole genome level, which has an increased demand on high-throughput and multiplexing methods that can be easily reprogram and repurpose DNA to generate genome-wide modifications. For these features CRISPR technology represent a new turning point for genome recoding and biological network rewiring, and it offers almost unlimited potentials in this direction. It has been not fully explored in *E. coli,* potentially because relatively efficient engineering methods were already available. Different form Eukaryotic cells, the CRISPR/Cas9 technology has had a major breakthrough in *E. coli* field when used as CRISPRi or CRISPR-TF rather than with its original endonuclease function. At this stage it may be difficult to imagine that CRISPR will substitute λ Red recombineering, which, in its simplicity, relies basically on electroporation of oligos and PCR products. Instead in order to introduce modification with CRISPR/Cas9 it necessary to always rely on a plasmid expressing the gRNA and one used as a donor, thus making this

system laborious and complex to just introduce one single modification and almost unfeasible to achieve multiplexing. However, CRISPR has recently been shown to be a very promising system if coupled with λ Red[65,106,107].

In this work we developed a system, CRMAGE, which drastically improves MAGE efficiency from 5% to 98% (**Chapter 7, Paper 1**). CRMAGE is based on the combination of MAGE recombineering and CRISPR/Cas9, which is used to induce targeted DSB to create a negative selection against the wild type sequence, in order to increase the recombineering performance. CRMAGE comes as a full, automatable and potentially high-throughput engineering platform that includes a web-based tool that facilitate the design of both the λ red oligo required for the specific mutation, as well as gRNA required for the negative selection. Since not all genomic targets are located in direct proximity to an PAM site, the software also enables the use of degenerated codons to broaden the range of sequence targets. CRMAGE is based on a user-cloning BioBrick system to speed up the design and cloning process. Using this two plasmids-based system we were able to achieve close to 98% efficiency for single point mutations versus 5 % using traditional MAGE, while 66% was reached for modulation of protein translation by replacing 6 bp RBS sequence compared to 6% using traditional MAGE. CRMAGE has furthermore been designed to enable efficient recycling of the plasmid used for negative selection and through the same approach it will also be possible to remove the entire

system in the last round of CRMAGE, where the CRISPR/Cas9 killing activity can target origin and antibiotics markers in both plasmids thus resulting in a clean recombinant strain. Moreover, the system enables simultaneous integration of multiple mutations with no apparent loss of efficiency. For multiplexing of CRMAGE targets, it is necessary to express multiple gRNAs. However, the standard design used for synthetic sgRNA have significant stretches of sequence homology (typically 136 nucleotides) that may result in homologous recombination[21]. For this reason, we have shown that the presence of a constitutively transcribed trascRNA that can be repurposed to generate multiple gRNAs from a CRISPR array. The synthesis and cloning of such an array of gRNAs (target-repeat-target-repeat etc.) is simple and it minimizes the risk of recombination in the vector. Thus, the expression of the pre-cRNA for multiplex CRMAGE may substitute the sgRNAs if put under control of the inducible pLtet promoter, making it easy to control the timing of the expression.

In conclusion, CRMAGE enables the generation of multiple mutations in a single cycle and multiple cycles within one working day, it has the potential to significantly increase the daily strain engineering capacity. The increased efficiency furthermore opens up the possibility of automating genome-scale engineering.

Chapter 4: Yeast engineering optimization

4.1 State of the art of *S. cerevisiae* engineering

Genetic manipulation in *S. cerevisiae* has always been quite easy due to the natural ability of this organism to recombine linear DNA in a very efficient way, as well as its efficient transformation capability. The use of Cre recombinase and FLP flippase[227–232] or other recombination based approaches, like the 50:50 method[233] enable marker recycling after each manipulations with knock-out or knock-in of exogenous DNA. The relatively high recombination efficiency of yeast has been also employed for combinatorial gene/pathway integration in δ sites with in vivo assembly of PCR product by using concatenated homologous arms [234,235].

Despite the amenability of these methods, they are relatively time consuming and therefore not suitable for the introduction of large heterologous metabolic pathways or deletion of several genes in a reasonably short time. In recent years new methods relying on the ability of double strand break to boost the repair system have been used to improve the efficiency of gene knock-out/in. These approaches typically use the customizable Zinc finger[127] and TAL activator-like effector coupled to nuclease domains[236] to induce a targeted DSB in order to promoter the recombination with the donor DNA at the defined cutting site. Although these nuclease-based systems are every effective, they are not suitable for multiplex simultaneous editing and they are not optimized for genome-scale

level approaches. For this reason DiCarlo *et al.* developed a mirrored system of the *E. coli* MAGE, the Yeast Oligo-Mediated Genome Engineering (YOGE)[237]. In their work they were able to achieve appreciable oligonucleotide insertion efficiencies frequencies of 0.2–2.0% without phenotypic selection, thus making it suitable for moderate screening efforts with an estimated library size of 10^2–10^5 recombinants per locus for each cycle.

Approaches like directed evolution with non-natural selective pressure combined with libraries of DNA parts derived by error-prone PCR or targeted mutagenesis have been used to identify specific traits, evolve synthetic devices and systems in order to expand the genomic diversity[238–240]. The advent of more high-throughput of DNA synthesis and assembly have brought yeast genetic to the level of a complete de novo synthesis of all 16 chromosomes[241], Sc2.0. In this effort, all nonessential genes will be flanked by *loxP* sites, allowing to SCRaMbLE the genome for random deletion of genes upon expression of Cre recombinase, combined with screening for viable strains with improved characteristics for a selectable trait[242,243].

4.2 Impact of CRISPR on *S. cerevisiae* engineering

CRISPR/Cas9 has been applied for targeted single and multiple gene deletions in *S. cerevisiae* by homology-directed repair of DSBs using short oligonucleotides as repair donors, in different strain backgrounds[98,100,244]. A pioneer work of CRISPR in yeast, carried out by DiCarlo *et al.*[98], demonstrated almost 100% recombination frequency without the use of a selectable marker by transforming linear double strand oligonucleotides as donor. This is a remarkably difference compared to YOGE where the integration is not mediated by DSB. The main advantages of endonuclease DSB mediated modifications over traditional techniques lay in their efficiency and accuracy since the inductions of DSB have been shown to increase the integration of heterologous DNA, as well as the accuracy of integration of heterologous linear DNA fragments with ends homologous to the DSB site[245,246]. Cas9 in particular appears as the most promising tool for rapid editing a genome at multiple loci.

Several approaches have demonstrated the efficiency of Cas9 by multiple integrations, knock-outs and in vivo assembly for in vivo pathway reconstruction and to generate quantitative gene assembly and DNA library insertion[99,101,102,244,247]. It has also been proven successful in the manipulation of strains that are difficult to engineer, including industrial strains[103]. Jakociunas *et al.* have demonstrated the powerful application of the CRISPR/Cas9 system as a tool for metabolic engineering utilizing user-

friendly and easy to use USER-technology-based gRNA constructs[247]. In order to further expand this existing platform with new engineering tools, we investigated whether CRISPR/Cas9 together with the DNA brick based EasyClone approach could be employed for targeted one-step selection-free integration of multiple genes into the *S. cerevisiae* genome (**Chapter 7**, **paper 2**)[104]. The basic idea of the work presented here was to use CRISPR/Cas9 for improving an already existing system, EasyClone, developed by Jensen *et al.*[248]. This EasyClone system consist of a set of vectors that enable fast and simultaneous multiple integrations of genes into specific "safe sites of insertion" with the possibility of recycling the selective markers[248]. The integration sites are located between essential elements, which limits the occurrence of chromosomal rearrangements due to the lethal effect this would cause[232]. Based on homologous recombination using 500 bp long homology arms, this method results in efficient integration into a single site. However, the efficiency of integration decreases when native genes or promoters are present on the fragment to be integrated, or in the case of multiple simultaneous integrations. Moreover, the method still relies on selective markers, and even though they are subsequently excided and recycled, this process requires more time and work. It additionally leaves scars scattered around the genome that can recombine and cause chromosomal aberrations. Increasing the efficiency of targeted integration without selection is therefore important for accelerating

and potentially automating the strain engineering process. For this reason we developed a marker-free multi gene and pathway integration platform, CrEdit (CRISPR/Cas9 mediated genome Editing) (**Chapter 7**, **paper 2**)[104]. In this system we combined the high specificity of CRISPR/Cas9 DSB mediated editing with the genome engineering tool EasyClone for achieving highly efficient and accurate marker-free simultaneous genomic integration of multiple pathway gene expression cassettes in different loci in the genome of *S. cerevisiae*. Using CrEdit we were able to achieve up to 100% correct selection-free target integrations at the desired locus and a very high efficient simultaneous integrations of three pathway genes involved in the production of β-carotene at three different integration sites located on three different chromosomes without applying any selection pressure. CrEdit therefore enables efficient genome editing without the inclusion of selection markers, thereby saving time and eliminating the generation of scars associated with their removal.

This system has proven to be a powerful tool for yeast engineering by eliminating the need of sequential introduction of gene modifications and marker recycle thus enabling marker-free and simultaneous multiple editing. Moreover, it offers the possibility to generate strains with combinatorial genetic modifications in just one transformation experiment.

Chapter 5: Revolution of CHO engineering

5.1 Unraveling Chinese Hamster Ovary (CHO) cells

Chinese hamster ovary (CHO) cells are the primary factories for biopharmaceuticals because of their capacity to correctly fold and post-translationally modify recombinant proteins compatible with humans[249]. CHO studies have been hindered by the inaccessibility to modification of the genome and thus to access the genetic information necessary for the optimization of production, until recently when the genome sequence of CHO-K1 and Chinese Hamster was revealed[250,251]. The significant number of technological advances focusing on phenotypic screening and growth media optimization[249,252] were achieved before the whole genome was sequenced. This has rendered CHO cells the workhorse of biopharmaceuticals, accounting for 35,5% of the total approved products in the market since 1987, when tissue plasminogen activator was first issued[249,252]. The unraveling of CHO-K1 genome sequence initiated a new -omics era for CHO biology. Studies on proteomics, genome-scale modeling, genomics, metabolomics, transcriptomics, glycomics and fluxomics have emerged in the past few years [253]. The availability of these huge datasets and the genome-scale models will enable a more rational design of CHO as cell factories like other organism have been already for decades such as *E. coli* or yeast. However the growing plethora of potential targets for increasing

cellular production capabilities and for improving product quality requires the development of new engineering technologies. The field of mammalian cell engineering technology has been developing quite slowly due to the difficulties in obtaining engineered cells with stable genome modifications. Antisense RNA[254–256], PNA (Peptide Nucelic Acid) [257] and other interference mechanisms have been extensively used as tools for modulation of gene expression by targeting translation or transcription processes thus they were especially engaged for high-throughput screening and for the characterization of gene function and regulation. However, these systems do not cause permanent genetic modifications but just transient phenotypes associated to gene knock down. Historically the allelic replacement and gene knock-outs were achieved by homologous recombination followed by selection although with very low efficiency [258]. Only in recent years break-through technologies such as meganulease[259,260], Zinc-Finger-phusion nuclease (ZFN)[127,261–264] and TALENs[128,265] have allowed researchers to modify and tailor specific DNA sequences creating stable genome modifications with very high efficiency[266]. The principles of CHO genome editing rely on the induction of cellular DNA repair systems upon DNA damage[267,268]. The dsDNA breaks introduced by programmable nucleases are repaired by either non-homologous end-joining (NHEJ) or homology-directed repair (HDR). The NHEJ is an error-prone mechanism, so enable to generate gene knock-outs due to the indels introduced at the target site

within the ORFs resulting in frameshift. HDR system instead allows specific allelic exchange and knock-ins but it does not occur frequently unless stimulated by the induced DSB of the nucleases[269–271]. However, the cost and amount of work required to produce tailored proteins is very high and therefore impacts on the speed and economy of large-scale projects. Additionally, it makes infeasible to perform high-throughput and genome-scale studies.

5.2 CHO engineering revolution with CRISPR technology

In the course of the past two years the introduction of CRISPR/Cas9 (Cluster Regularly Interspaced Short Palindromic Repeats) genome editing technology in mammalian cells such as human and other cells lines[69–71,74,77,80,121,125,272] has circumvented these problems by decreasing the cost and time of cell line construction and allowing studies at genome-scale level[68]. Indeed, besides introducing single of few simultaneous mutations at the time with high efficiency [69,70,74,75,77,78,273], several studies have shown the effective application of this system for multiplexing at the whole genome-scale level by exploring phenotypic variation associated to desired phenotypes or in more ambitious cases to study cancer genesis and cancer markers genotyping[73,123–126,274]. Moreover, by introducing multiple simultaneous DSB it is possible to cause large kilo-base or mega-base deletions or even entire chromosomal rearrangements[77,275–277], which makes it possible to investigate the effect of whole gene cluster deletions or the characterization of

phenotypes associated to structural changes of the chromosome and aneuploidy state of human cells[276,277].

The ability of Cas9 to create targeted DSBs has also been used to boost the HR efficiency resulting in high efficiency of DNA insertions by co-transfecting either plasmid DNA or single stranded donor oligonucleotides (ssODNs) with 40–50 bp of short flanking homologous sequences [76,121,270,278]. Additionally, high efficiency of insertion could be achieved also via non HDR-mechanism as microhomology-mediated end joining (MMEJ) and NHEJ repairs[119,279]. The use of the nickase, as already mentioned in Chapter 2, have been used to increase the stringency[70,119–121]. The use of CRISPRi and CRISPR-TFs have also improved the understanding of cell physiology and they have been applied for the investigation of gene functions as well as rewiring regulatory networks as previously discussed in Chapter 2.

Our study (**Chapter 7**, **paper 3**) has been one of the pioneers in the field, being the first demonstrating the application of a CHO codon-optimized Cas9 for CHO genome editing by disrupting COSMC and FUT8, genes coding for proteins involved in O- and N- glycosylation, respectively [280,281]. Additionally, we developed a web-based bioinformatics tool "CRISPy" to assist in the design of sgRNAs and to predict possible off-targets based on the annotated CHO-K1 genome in order to facilitate high-throughput automated gene disruptions in CHO. Therefore we demonstrated the high efficiency (we were able to achieve up to 47,3% of indels efficiency without

selection), robustness, ease of use, and low costs of CRISPR/Cas9 system in CHO, thus showing the potential of this system as a genome-editing tool for both the academic and industrial community. Moreover, the introduction of the CRISPR/Cas9 system in CHO cells combined with the CRISPy design tool will contribute to improve the genome editing tool box and cell network-interactions understanding in CHO cells and enhance will contribute to improve the platform of biopharmaceuticals production in terms of yields and quality. This technology in mammalian cells has been mainly used with the intention of screening for gene function rather than applied to engineer cells for cell factories. We instead have been among the first to foresee its industrial potential by filling a patent (**Chapter 7**, **EPO Patent Application WO2015052231**) on the development of a multiplex editing system, especially suitable for mammalian cells, where a cell line with stable integrated and inducible Cas9 is used as engineering cell line for screening and fast development of new cell lines optimized for production. We have integrated the cell line with a system that enables virtually infinite integrations of the desire genes or pathways in safe "landing harbors" by regenerating the target sites (Fig. 4), beside the possibility of multiplexing knocks out and knock- ins in correlation with the amount of induction (Figs. 5-6).

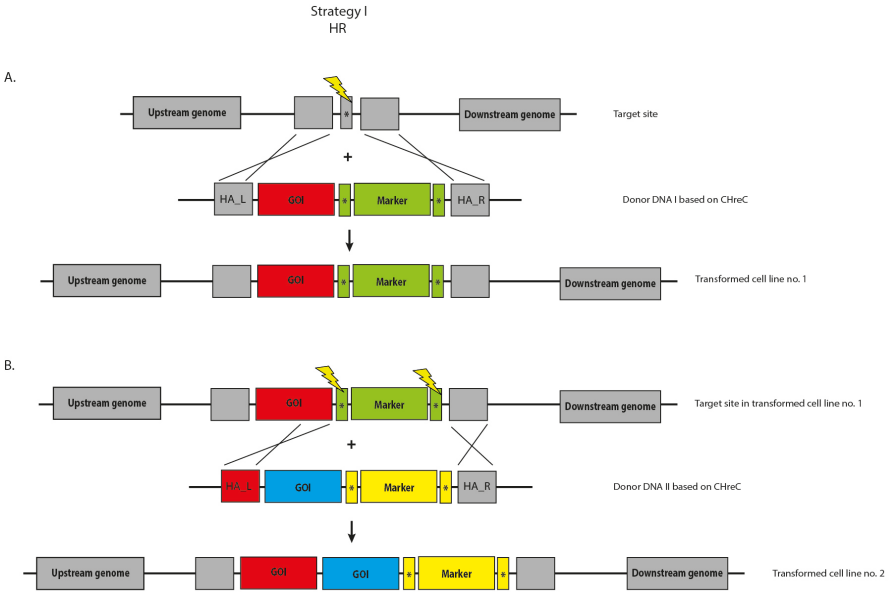


Fig. 4: **The CRESC insert strategy using homologous recombination**. Figure adapted from C.Ronda, L.E. Pedersen, H.F. Kildegaard, J.S. Lee. ***"Multiplex editing system***". EPO Patent Application (WO2015052231).

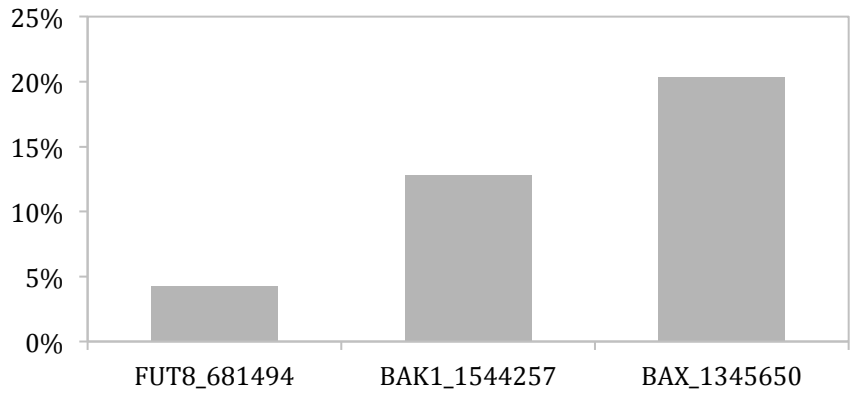


Fig. 5: Percentage of indels generated in 2AGFP_Cas9 cell line. Figure adapted from C.Ronda, L.E. Pedersen, H.F. Kildegaard, J.S. Lee. ***"Multiplex editing system***". EPO Patent Application (WO2015052231). Preliminary data from polyclonal FACS enriched population and not single clone population with optimized Cas9 expression. The percentage of indels created at FUT8, BAK and BAX loci upon induction was analyzed by MiSeq sequencing of target loci.

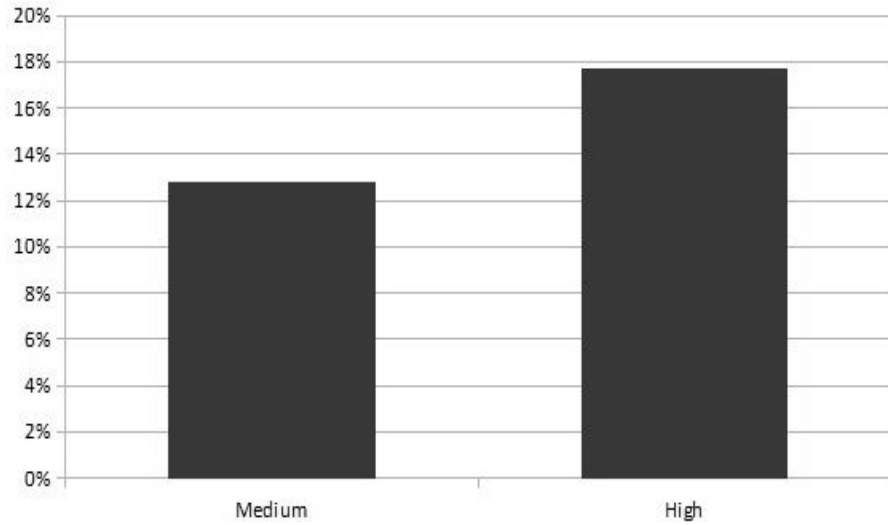


Fig. 6: Percentage of indels generated from permanently integrated Cas9 sorted for medium of high GFP expression. The percentage of indels created at FUT8 loci upon treatment of CHO cells with permanently integrated Cas9 and transiently transfected guideRNA was analyzed by MiSeq sequencing of target loci after FACS sorting for GFP.

Chapter 6: Conclusions and future prospective

CRISRP based technology has opened a completely new branch of research due to its incredible pliability to any type of organism and to its very diverse applications. The incredible ability to reprogram a protein towards any genetic sequence has made Cas9 the Holy Grail for synthetic biologists since it can be fused to any catalytic domain, and thereby reassigning a new function, while still maintaining the sequence specific and programmable nature. The accuracy, efficiency, and cost of CRISPR/Cas9 makes it an attractive alternative to other methods of genome modification and to approach organisms that are difficult to engineer. This enables multiplexing and the creation of high-throughput libraries for screening, designing of complex circuit to reprogram cell behavior and redirecting cell fluxes or to be used as scaffold for imaging platforms. Even split Cas9 has emerged for new applications which are still to be developed[282]. Despite the great achievements, there are still caveats that must be addressed in order to improve and optimize its function and for expanding the engineering toolbox. Indeed the possible off-target effects still raise concern and need to be addressed once and for all before considering to use the technology therapeutically. This is also important to avoid unforeseen pleiotropic effects during genome-scale experiments. Additionally, the PAM sequence, albeit very frequent, is a considerable constrain for the target range. Moreover, in

other organisms like *S. cerevisiae* and bacteria such as *E. coli*, the ability of generating huge screening library is not an option since the NHEJ system is lacking or is not preferentially used and therefore DSB are usually repaired by HDR that requires a donor DNA and no indels are created. Besides that, gRNAs need to be expressed through plasmids, and in this way the system reaches a bottleneck when multiple gRNAs have to be expressed by one plasmid. For *E. coli* it is even more challenging since the multiple donors DNA required have to be transformed as plasmids since linearized DNA is degraded immediately.

In conclusion even though CRISPR technology offers great prospective for next generation of cell factories, basic research and pharmaceutical industry (and there is still a lot to come with the discovery of new CRISPR systems or the optimization of CRISPR/Cas9), a moratorium on human germline modification has been called for. A huge ethical debate has begun focusing on the possibility of giving to the humankind the power to control its own evolution at the genome level[283]. Is eugenics and uncontrolled generation of GMO organisms lurking behind CRISPR technology?

Bibliography

1. Crane, H., Kinderman, E. & Malhotra, R. *A Cubic Mile of Oil: Realities and Options for Averting the Looming Global Energy Crisis*. (Oxford University Press, 2010).

2. Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* **4,** 259–288 (2013).

3. Kavšček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. Yeast as a cell factory: current state and perspectives. *Microb. Cell Fact.* **14,** 94 (2015).

4. Bailey, J. E. Toward a science of metabolic engineering. *Science.* **252,** 1668–1675 (1991).

5. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8,** 172 (2007).

6. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).

7. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

8. Makarova, K. S., Aravind, L., Grishin, N. V, Rogozin, I. B. & Koonin, E. V. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30,** 482–496 (2002).

9. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

10. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).

11. Makarova, K. S. *et al.* Evolution and classification of the CRISPR/Cas systems. *Nat. Rev. Microbiol.* **9,** 467–477 (2011).

12. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science.* **315,** 1709–1712 (2007).

13. Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8,** R61 (2007).

14. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comput. Biol.* **1,** e60 (2005).

15. Makarova, K. S., Grishin, N. V, Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1,** 7 (2006).

16. Makarova, K. S. & Koonin, E. V. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol. Biol.* **1311,** 47–75 (2015).

17. Heidrich, N. & Vogel, J. Same same but different: new structural insight into CRISPR-Cas complexes. *Mol. Cell* **52,** 4–7 (2013).

18. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42,** 2577–2590 (2014).

19. Godde, J. S. & Bickerton, A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **62,** 718–729 (2006).

20. Lillestøl, R. K. *et al.* CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72,** 259–272 (2009).

21. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494,** 489–491 (2013).

22. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21,** 1616–1625 (2011).

23. Minot, S., Bryson, A. & Chehoud, C. Rapid evolution of the human gut virome. in *Proc. Natl. Acad. Sci.* (2013).

24. Van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12,** 479–492 (2014).

25. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

26. Sontheimer, E. J. & Marraffini, L. A. Microbiology: slicer for DNA. *Nature* **468,** 45–46 (2010).

27. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).

28. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10,** 200–207 (2008).

29. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–1400 (2008).

30. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).

31. Shah, S. A., Erdmann, S., Mojica, F. J. M. & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10,** 891–899 (2013).

32. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7,** e35888 (2012).

33. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40,** 5569–5576 (2012).

34. Erdmann, S. & Garrett, R. A. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol. Microbiol.* **85,** 1044–1056 (2012).

35. Stern, A., Keren, L., Wurtzel, O., Amitai, G. & Sorek, R. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* **26,** 335–340 (2010).

36. Vercoe, R. B. *et al.* Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* **9,** e1003454 (2013).

37. Reeks, J., Naismith, J. H. & White, M. F. CRISPR interference: a structural perspective. *Biochem. J.* **453,** 155–166 (2013).

38. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes {&} Dev.* **22,** 3489–3496 (2008).

39. Garside, E. L. *et al.* Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA* **18,** 2020–2028 (2012).

40. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18,** 529–536 (2011).

41. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci.* **108,** 10092–10097 (2011).

42. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* . **321,** 960–964 (2008).

43. Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139,** 945–956 (2009).

44. Hatoum-Aslan, A., Maniv, I. & Marraffini, L. A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci.* **108,** 21218–21222 (2011).

45. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).

46. Westra, E. R. *et al.* The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.* **46,** 311–339 (2012).

47. Westra, E. R. *et al.* Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biol.* **9,** 1134–1138 (2012).

48. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46,** 606–615 (2012).

49. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci.* **108,** 10098–10103 (2011).

50. Wiedenheft, B. *et al.* Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477,** 486–489 (2011).

51. Spilman, M. *et al.* Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol. Cell* **52,** 146–152 (2013).

52. Hochstrasser, M. L. *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci.* **111,** 6618–6623 (2014).

53. Westra, E. R. *et al.* CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* **46,** 595–605 (2012).

54. Westra, E. R. *et al.* Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.* **9,** e1003742 (2013).

55. Gasiunas, G. *et al.* Cas9crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. in *Proc. Natl. Acad. Sci.* **109,** E2579–E2586 (2012).

56. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337,** 816–821 (2012).

57. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3,** 945 (2012).

58. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–7 (2014).

59. Almendros, C., Guzmán, N. M., Díez-Villaseñor, C., García-Martínez, J. & Mojica, F. J. M. Target Motifs Affecting Natural Immunity by a Constitutive CRISPR-Cas System in Escherichia coli. *PLoS One* **7,** e50797 (2012).

60. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463,** 568–571 (2010).

61. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

62. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science.* **343,** 1247997 (2014).

63. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

64. Magadán, A. H., Dupuis, M. È., Villion, M. & Moineau, S. Cleavage of phage DNA by the Streptococcus thermophilus CRISPR3-Cas system. *PLoS One* **7,** (2012).

65. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

66. Koo, Y., Ka, D., Kim, E. J., Suh, N. & Bae, E. Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *J. Mol. Biol.* **425,** 3799–3810 (2013).

67. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. PNAS Plus: Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109,** E2579–E2586 (2012).

68. Pennisi, E. The CRISPR Craze. *Science.* **341,** 833–836 (2013).

69. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science.* **339,** 823–826 (2013).

70. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science.* **339,** 819–823 (2013).

71. Xue, H., Wu, J., Li, S., Rao, M. S. & Liu, Y. Genetic Modification in Human Pluripotent Stem Cells by Homologous Recombination and CRISPR/Cas9 System. *Methods Mol. Biol.* (2014). doi:10.1007/7651_2014_73

72. Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* **41,** e188–e188 (2013).

73. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–7 (2014).

74. Ronda, C. *et al.* Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* **111,** 1604–1616 (2014).

75. Grav, L. M. *et al.* One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* (2015).

76. Lee, J. S., Kallehauge, T. B., Pedersen, L. E. & Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* **5,** 8572 (2015).

77. Canver, M. C. *et al.* Characterization of Genomic Deletion Efficiency Mediated by CRISPR/Cas9 in Mammalian Cells. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M114.564625

78. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–918 (2013).

79. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell* **154,** 1370–1379 (2013).

80. Yasue, A. *et al.* Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems. *Sci. Rep.* **4,** 5705 (2014).

81. Yen, S. T. *et al.* Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. *Dev. Biol.* **393,** 3–9 (2014).

82. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep.* **4,** 220–228 (2013).

83. Gratz, S. J. *et al.* Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196,** 961–971 (2014).

84. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc. Natl. Acad. Sci.* E2967–E2076 (2014). doi:10.1073/pnas.1405500111

85. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res* **23,** 465–472 (2013).

86. Hwang WY *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 227–229 (2013).

87. Jao, L.-E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci.* **110,** 13904–9 (2013).

88. Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24,** 142–153 (2014).

89. Krinsky, N. I., Landrum, J. T. & Bone, R. A. Biologic mechanisms of the protective role of lutein and zeaxanthin in the eye. *Annu. Rev. Nutr.* **23,** 171–201 (2003).

90. Hruscha, A. *et al.* Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140,** 4982–7 (2013).

91. Blitz, I. L., Biesinger, J., Xie, X. & Cho, K. W. Y. Biallelic genome modification in F0 Xenopus tropicalis embryos using the CRISPR/Cas system. *Genesis* **51,** 827–834 (2013).

92. Friedland, A. E. *et al.* Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat. Methods* **10,** 741–3 (2013).

93. Wang, Y. *et al.* The CRISPR/Cas system mediates efficient genome engineering in Bombyx mori. *Cell Res.* **23,** 1414–1416 (2013).

94. Li, J.-F. *et al.* Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnol.* **31,** 686–688 (2013).

95. Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 686–688 (2013).

96. Belhaj, K., Chaparro-Garcia, A., Kamoun, S. & Nekrasov, V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* **9,** 39 (2013).

97. Niu, Y. *et al.* Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156,** 836–843 (2014).

98. DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41,** 4336–4343 (2013).

99. Mans, R. *et al.* CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in

Saccharomyces cerevisiae. *FEMS Yeast Res.* **15,** fov004–fov004 (2015).

100. Bao, Z. *et al.* Homology-Integrated CRISPR{\textendash}Cas (HI-CRISPR) System for One-Step Multigene Disruption in Saccharomyces cerevisiae. *ACS Synth. Biol.* (2014). at <doi/abs/10.1021/sb500255k>

101. Horwitz, A. A., Walter, J. M., Schubert, M. G. & Kung, S. H. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst.* (2015).

102. Jakočiunas, T. *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metab. Eng.* (2015).

103. Stovicek, V., Borodina, I. & Forster, J. CRISPR-Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains. *Metab. Eng. Commun.* **2,** 13–22 (2015).

104. Ronda, C. *et al.* CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. *Microb. Cell Fact.* **14,** 97 (2015).

105. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* (2015).

106. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).

107. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system to lambda Red recombineering enables simplified chromosomal gene replacement in Escherichia coli. *Appl. Environ. Microbiol.* (2015).

108. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2015).

109. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 231–243 (2015).

110. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–6 (2013).

111. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–32 (2013).

112. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** 839–43 (2013).

113. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32,** 670–676 (2014).

114. Cencic, R. *et al.* Protospacer Adjacent Motif (PAM)-Distal Sequences Engage CRISPR Cas9 DNA Target Cleavage. *PLoS One* **9,** e109213 (2014).

115. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–198 (2014).

116. Wang, X. *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase- defective lentiviral vectors. *Nat Biotechnol* **33,** (2015).

117. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12,** 237–244 (2015).

118. Fujii, W., Kawasaki, K., Sugiura, K. & Naito, K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res.* **41,** (2013).

119. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154,** 1380–1389 (2013).

120. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11,** 399–402 (2014).

121. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–308 (2013).

122. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32,** 577–82 (2014).

123. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–4 (2014).

124. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32,** 267–73 (2014).

125. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509,** 487–91 (2014).

126. Chen, S. *et al.* Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis Resource Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell Res* 1246–1260 (2015).

127. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188,** 773–782 (2011).

128. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49–55 (2013).

129. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* **3,** 737–747 (2002).

130. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21,** 635–637 (2003).

131. Gomaa, A. A. *et al.* Programmable removal of bacterial strains by use of genome- targeting CRISPR-cas systems. *MBio* **5,** (2014).

132. Citorik, R. J., Mimee, M. & Lu, T. K. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* **32,** (2014).

133. Bikard, D. *et al.* Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* **32,** 1141–1145 (2014).

134. Yosef, I., Manor, M., Kiro, R. & Qimron, U. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci.* **2015,** 201500107 (2015).

135. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).

136. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196 (2013).

137. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437 (2013).

138. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–51 (2013).

139. Nielsen, A. A. K. & Voigt, C. A. Multi-input CRISPR / Cas genetic circuits that interface host regulatory networks. *Mol. Syst. Biol.* 1–11 (2014).

140. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10,** 973–6 (2013).

141. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* **2,** 604–613 (2013).

142. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54,** 698–710 (2014).

143. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell Res* **9,** 339–350 (2015).

144. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 1–18 (2014). doi:10.1038/nature14136

145. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell Res* 647–661 (2014).

146. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–21 (2013).

147. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** (2015).

148. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11,** 198–200 (2015).

149. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based Photoactivatable Transcription System. *Chem. Biol.* **22,** 169–174 (2015).

150. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR / Cas System. *Cell* **155,** 1479–1491 (2013).

151. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159,** 635–646 (2014).

152. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci.* (2015). doi:10.1073/pnas.1420024112/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1420024112

153. Haurwitz, R. E. *et al.* Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science.* **329,** 1355–1358 (2010).

154. Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. & Arkin, A. P. RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30,** 1002–1006 (2012).

155. Lee, H. Y. *et al.* RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci.* **110,** 5416–5421 (2013).

156. Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43,** 674–81 (2015).

157. Feist, A. M. & Palsson, B. Ø. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat. Biotechnol.* **26,** 659–667 (2008).

158. Feist, A. M. *et al.* Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab. Eng.* **12,** 173–186 (2010).

159. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6,** 1290–1307 (2011).

160. Chen, X. *et al.* Metabolic engineering of Escherichia coli: a sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* **31,** 1200–1223 (2013).

161. Thomason, L. C., Costantino, N. & Court, D. L. E. coli genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **Chapter 1,** Unit 1.17 (2007).

162. Karberg, M. *et al.* Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.* **19,** 1162–1167 (2001).

163. Yao, J. *et al.* Use of targetrons to disrupt essential and nonessential genes in Staphylococcus aureus reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA* **12,** 1271–1281 (2006).

164. Shao, L. *et al.* Targeted gene disruption by use of a group II intron (targetron) vector in Clostridium acetobutylicum. *Cell Res.* **17,** 963–965 (2007).

165. Enyeart, P. J. *et al.* Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol. Syst. Biol.* **9,** 685 (2013).

166. Enyeart, P. J., Mohr, G., Ellington, A. D. & Lambowitz, A. M. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mob. DNA* **5,** 1–19 (2014).

167. Perutka, J., Wang, W., Goerlitz, D. & Lambowitz, A. M. Use of Computer-designed Group II Introns to Disrupt Escherichia coli DExH/D-box Protein and DNA Helicase Genes. *J. Mol. Biol.* **336,** 421–439 (2004).

168. Guo, H. *et al.* Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289,** 452–457 (2000).

169. Kilby, N. J., Snaith, M. R. & Murray, J. A. Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9,** 413–421 (1993).

170. Yu, B. J. & Kim, C. Minimization of the Escherichia coli genome using the Tn5-targeted Cre/loxP excision system. *Methods Mol. Biol.* **416,** 261–277 (2008).

171. Delneri, D. *et al.* Exploring redundancy in the yeast genome: An improved strategy for use of the cre-loxP system. *Gene* **252,** 127–135 (2000).

172. Suzuki, N., Tsuge, Y., Inui, M. & Yukawa, H. Cre/loxP-mediated deletion system for large genome rearrangements in Corynebacterium glutamicum. *Appl. Microbiol. Biotechnol.* **67,** 225–233 (2005).

173. Campo, N., Daveran-Mingot, M. L., Leenhouts, K., Ritzenthaler, P. & Le Bourgeois, P. Cre-loxP recombination system for large genome rearrangements in Lactococcus lactis. *Appl. Environ. Microbiol.* **68,** 2359–2367 (2002).

174. Court, D. L., Sawitzke, J. A. & Thomason, L. C. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36,** 361–388 (2002).

175. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–6645 (2000).

176. Muyrers, J. P. P., Zhang, Y., Buchholz, F. & Stewart, A. F. RecE/RecT and Redα/Redβ initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* **14,** 1971–1982 (2000).

177. Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci.* **97,** 5978–5983 (2000).

178. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4,** 206–223 (2009).

179. Sukhija, K. *et al.* Developing an extended genomic engineering approach based on recombineering to knock-in heterologous genes to Escherichia coli genome. *Mol. Biotechnol.* **51,** 109–118 (2012).

180. Bird, A. W. *et al.* High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat. Methods* **9,** 103–109 (2011).

181. Zhang, Y., Muyrers, J. P., Testa, G. & Stewart, A. F. DNA cloning by homologous recombination in Escherichia coli. *Nat. Biotechnol.* **18,** 1314–1317 (2000).

182. Murphy, K. C., Campellone, K. G. & Poteete, A. R. PCR-mediated gene replacement in Escherichia coli. *Gene* **246,** 321–330 (2000).

183. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci.* **98,** 6742–6746 (2001).

184. Swaminathan, S. *et al.* Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29,** 14–21 (2001).

185. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27,** 465–71 (2009).

186. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20,** 123–128 (1998).

187. Kmiec, E. & Holloman, W. K. Beta protein of bacteriophage lambda promotes renaturation of DNA. *J. Biol. Chem.* **256,** 12636–12639 (1981).

188. Li, Z., Karakousis, G., Chiu, S. K., Reddy, G. & Radding, C. M. The beta protein of phage lambda promotes strand exchange. *J. Mol. Biol.* **276,** 733–744 (1998).

189. Rybalchenko, N., Golub, E. I., Bi, B. & Radding, C. M. Strand invasion promoted by recombination protein beta of coliphage lambda. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 17056–17060 (2004).

190. Hall, S. D. & Kolodner, R. D. Homologous pairing and strand exchange promoted by the Escherichia coli RecT protein. *Proc. Natl. Acad. Sci.* **91,** 3205–3209 (1994).

191. Maresca, M. *et al.* Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.* **11,** 54 (2010).

192. Murphy, K. C. The Gam Protein Inhibits RecBCD Binding to dsDNA Ends. *J. Mol. Biol.* **371,** 19–24 (2007).

193. Marsic, N., Roje, S., Stojiljkovic, I., Salaj-Smic, E. & Trgovcevic, Z. In vivo studies on the interaction of RecBCD enzyme and lambda gam protein. *J. Bacteriol.* **175,** 4738–4743 (1993).

194. Iyer, L. M., Koonin, E. V & Aravind, L. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3,** 8 (2002).

195. Thomason, L. *et al.* Recombineering: genetic engineering in bacteria using homologous recombination. *Curr. Protoc. Mol. Biol.* **Chapter 1,** Unit 1.16 (2007).

196. Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2,** 769–779 (2001).

197. Lee, E. C. *et al.* A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73,** 56–65 (2001).

198. Oppenheim, A. B., Rattray, A. J., Bubunenko, M., Thomason, L. C. & Court, D. L. In vivo recombineering of bacteriophage λ by PCR fragments and single-strand oligonucleotides. *Virology* **319,** 185–189 (2004).

199. Blank, K., Hensel, M. & Gerlach, R. G. Rapid and Highly Efficient Method for Scarless Mutagenesis within the Salmonella enterica Chromosome. *PLoS One* **6,** e15763 (2011).

200. Yu, B. J. *et al.* Rapid and efficient construction of markerless deletions in the Escherichia coli genome. *Nucleic Acids Res.* **36,** e84–e84 (2008).

201. Yang, J. *et al.* High-efficiency scarless genetic modification in Escherichia coli by using lambda red recombination and I-SceI cleavage. *Appl. Environ. Microbiol.* **80,** 3826–3834 (2014).

202. Kim, J., Webb, A. M., Kershner, J. P., Blaskowski, S. & Copley, S. D. A versatile and highly efficient method for scarless genome editing in Escherichia coli and Salmonella enterica. *BMC Biotechnol.* **14,** 84 (2014).

203. Posfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F. R. Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* **27,** 4409–4415 (1999).

204. Benzinger, R., Enquist, L. W. & Skalka, A. Transfection of Escherichia coli spheroplasts. V. Activity of recBC nuclease in rec+ and rec minus spheroplasts measured with different forms of bacteriophage DNA. *J. Virol.* **15,** 861–871 (1975).

205. Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci.* **100,** 15748–15753 (2003).

206. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186,** 791–799 (2010).

207. Sawitzke, J. A. *et al.* Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol.* **407,** 45–59 (2011).

208. Oh, J. H. & van Pijkeren, J. P. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic Acids Res.* **42,** e131–e131 (2014).

209. Kosuri, S. *et al.* Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28,** 1295–1299 (2010).

210. Conrad, T. M., Lewis, N. E. & Palsson, B. Ø. Microbial laboratory evolution in the era of genome-scale science. *Mol. Syst. Biol.* **7,** 509 (2011).

211. Esvelt, K. M. & Wang, H. H. Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* **9,** 641 (2013).

212. Carr, P. A. & Church, G. M. Genome engineering. *Nat. Biotechnol.* **27,** 1151–1162 (2009).

213. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc. Natl. Acad. Sci.* **102,** 15971–15976 (2005).

214. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–60 (2013).

215. Lartigue, C. *et al.* Genome transplantation in bacteria: changing one species to another. *Science* **317,** 632–638 (2007).

216. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1,** 2005.0018 (2005).

217. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56 (2010).

218. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319,** 1215–1220 (2008).

219. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R. & Reznikoff, W. S. *Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat. Biotechnol.* **18,** (2000).

220. Winterberg, K. M., Luecke, J., Bruegl, A. S. & Reznikoff, W. S. Phenotypic screening of Eschenchia coli K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays. *Appl. Environ. Microbiol.* **71,** 451–459 (2005).

221. Pósfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312,** 1044–1046 (2006).

222. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).

223. Wang, H. H. & Church, G. M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol.* **498,** 409–426 (2011).

224. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28,** 856–862 (2010).

225. Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333,** 348–353 (2011).

226. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* **40,** e132–e132 (2012).

227. Sauer, B. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.* **7,** 2087–2096 (1987).

228. Sauer, B. Multiplex Cre/lox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. *Nucleic Acids Res.* **24,** 4608–4613 (1996).

229. Yoon, Y. G., Pósfai, G., Szybalski, W. & Kim, S. C. Cre/loxP-mediated in vivo excision of large segments from yeast genome and their amplification based on the 2 μm plasmid-derived system. in *Gene* **223,** 67–76 (1998).

230. Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D. & Hegemann, J. H. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30,** e23 (2002).

231. Radhakrishnan, P. & Srivastava, V. Utility of the FLP-FRT recombination system for genetic manipulation of rice. *Plant Cell Rep.* **23,** 721–726 (2005).

232. Mikkelsen, M. D. *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14,** 104–111 (2012).

233. Horecka, J. & Davis, R. W. The 50:50 method for PCR-based seamless genome editing in yeast. *Yeast* (2013). doi:10.1002/yea.2992

234. Yuan, J. & Ching, C. B. Combinatorial Assembly of Large Biochemical Pathways into Yeast Chromosomes for Improved Production of Value-added Compounds. *ACS Synth. Biol.* (2014). doi:10.1021/sb500079f

235. Shao, Z., Zhao, H. H. & Zhao, H. H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37,** e16–e16 (2009).

236. Bogdanove, A. J. & Voytas, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science.* **333,** 1843–1846 (2011).

237. Dicarlo, J. E. *et al.* Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.* **2,** 741–749 (2013).

238. Tyo, K. E. J., Nevoigt, E. & Stephanopoulos, G. Directed evolution of promoters and tandem gene arrays for customizing RNA synthesis rates and regulation. *Methods Enzymol.* **497,** 135–155 (2011).

239. Rajkumar, A. S., Dénervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nat. Genet.* **45,** 1207–15 (2013).

240. Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32,** 97–101 (2014).

241. Annaluru, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science.* **344,** 55–58 (2014).

242. Dymond, J. & Boeke, J. The saccharomyces cerevisiae SCRaMbLE system and genome minimization. *Bioeng. Bugs* **3,** 168–171 (2012).

243. Jovicevic, D., Blount, B. A. & Ellis, T. Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. *BioEssays* **36,** 855–860 (2014).

244. Ryan, O. W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* **3,** (2014).

245. Storici, F., Durham, C. L., Gordenin, D. A. & Resnick, M. A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci.* **100,** 14994–14999 (2003).

246. Kuijpers, N. G. A. *et al.* A versatile, efficient strategy for assembly of multi-fragment expression vectors in Saccharomyces cerevisiae using 60 bp synthetic recombination sequences. *Microb. Cell Fact.* **12,** 47 (2013).

247. Jakociunas, T. *et al.* CasEMBLR: Cas9-facilitated multi-loci genomic integration of in vivo assembled DNA parts in Saccharomyces cerevisiae. *ACS Synth. Biol.* (2015). doi:10.1021/acssynbio.5b00007

248. Jensen, N. B. *et al.* EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae. *FEMS Yeast Res.* **14,** 238–248 (2014).

249. Jayapal, K., Wlaschin, K., Hu, W. & Yap, G. Recombinant protein therapeutics from CHO cells-20 years and counting. *Chem. Eng. Prog.* **103,** 40–47 (2007).

250. Xu, X. *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29,** 735–741 (2011).

251. Lewis, N. E. *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* **31,** 759–65 (2013).

252. Walsh, G. Biopharmaceutical_benchmarks_2014. **32,** 1–10 (2014).

253. Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E. & Betenbaugh, M. J. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* **24,** 1102–1107 (2013).

254. Mohr, S. E. & Perrimon, N. RNAi screening: New approaches, understandings, and organisms. *Wiley Interdiscip. Rev. RNA* **3,** 145–158 (2012).

255. Lim, S. F. *et al.* RNAi suppression of Bax and Bak enhances viability in fed-batch cultures of CHO cells. *Metab. Eng.* **8,** 509–522 (2006).

256. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat. Rev. Genet.* **7,** 373–384 (2006).

257. Shakeel, S., Karim, S. & Ali, A. Peptide nucleic acid (PNA) - A review. *J. Chem. Technol. Biotechnol.* **81,** 892–899 (2006).

258. Yamane-Ohnuki, N. *et al.* Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol. Bioeng.* **87,** 614–622 (2004).

259. Galetto, R., Duchateau, P. & Pâques, F. Targeted approaches for gene therapy and the emergence of engineered meganucleases. *Expert Opin. Biol. Ther.* **9,** 1289–1303 (2009).

260. Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.* **11,** 11–27 (2011).

261. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci.* **104,** 3055–3060 (2007).

262. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435,** 646–651 (2005).

263. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8,** 753–755 (2011).

264. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci.* **105,** 5809–5814 (2008).

265. Miller JC *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–148 (2011).

266. Carroll, D. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* **83,** 409–39 (2014).

267. Gaj, T., Gersbach, C. a & Iii, C. F. B. ZFN , TALEN , and CRISPR / Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).

268. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–34 (2014).

269. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).

270. Maggio, I. & Gonçalves, M. A. F. V. Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends Biotechnol.* **33,** 280–291 (2015).

271. Sedivy, J. M. & Sharp, P. A. Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proc. Natl. Acad. Sci.* **86,** 227–231 (1989).

272. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 230–2 (2013).

273. Wilkinson, R. & Wiedenheft, B. A CRISPR method for genome engineering. *F1000Prime Rep.* **6,** 3 (2014).

274. Kim, J. M., Kim, D., Kim, S. & Kim, J.-S. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat. Commun.* **5,** 3157 (2014).

275. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR Cas-derived RNA-guided endonucleases and nickases sup2. 132–141 (2014). doi:10.1101/gr.162339.113.Freely

276. Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5,** 3728 (2014).

277. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. 2059–2065 (2014). doi:10.1101/gr.177220.114.

278. Perez-Pinera, P., Ousterout, D. G. & Gersbach, C. A. Advances in targeted genome editing. *Curr. Opin. Chem. Biol.* **16,** 268–277 (2012).

279. Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* 1–3 (2014). doi:10.1038/ncomms6560

280. Miyoshi, E. *et al.* The alpha1-6-fucosyltransferase gene and its biological significance. *Biochim. Biophys. Acta* **1473,** 9–20 (1999).

281. Wang, Y. *et al.* Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc. Natl. Acad. Sci.* **107,** 9228–9233 (2010).

282. Wright, A. V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci.* **112,** 2984–2989 (2015).

283. E. Lanphier, F, Urnov, Fyodor, S.E Haecker, M. Werner, J. Smolenski. Don ' t edit the human germ line. *Nat.* **519,** 410-1(2015)

Chapter 7: Full length papers and patent

Paper 1.

Ronda C., L.E. Pedersen, A.T. Nielsen. **CRMAGE: CRISPR Optimized MAGE Recombineering combined using web-based oligos design.** *Scie.Rep.* Submitted.

CRMAGE: CRISPR Optimized MAGE Recombineering combined using web-based oligo design

Carlotta Ronda[1,1], Lasse Ebdrup Pedersen[1,2], Alex Toftgaard Nielsen[1,3#]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark

[1]carro@biosustain.dtu.dk, [2]laeb@biosustain.dtu.dk, [3]atn@biosustain.dtu.dk

[#] Corresponding author

Phone: +45 45258010

Postal Address: Kogle Allé 6, DK-2970 Hørsholm, Denmark

Email: atn@biosustain.dtu.dk

Abstract

Metabolic engineering is currently pushing forward new frontiers by using combinations of metabolic network reconstructions, efficient laboratory evolution, and genome sequencing methods. One of the current bottlenecks is efficient genome engineering technologies required for creating the desired genotypes. Here, we combine CRISPR and MAGE technology (CRMAGE) to create a highly efficient and fast method for genome engineering of *E. coli*. This two plasmids based system combines λ red recombineering and CRISPR/Cas9. Using CRMAGE we were able to increase MAGE performance from 5% to 97.6% efficiency for gene recoding that leads to a truncated protein and from 6% to between 62% to 70% efficiency for modulation of protein synthesis (small insertion/RBS substitution). The system also enables introduction of at least two mutations in a single round of recombineering with similar high efficiencies. Degenerate codon usage can be used to expand the range of CRISPR targets, thus creating a powerful tool for targeting any site in the genome. The CRMAGE system is based two plasmids that use a USER-cloning-assembly-based system enabling quick and cost efficient gRNA target replacement. The system furthermore utilizes a CRISPR system targeting the plasmids, thereby enabling fast recycling of the vectors. A web-based tool was furthermore developed to predict a customized λ red oligo required for the mutation of the desired nucleotides in the genome, as well as the possible gRNAs that can be used in combination with it. In conclusion, CRMAGE platform enables highly efficient and fast genome editing by combining MAGE with recombineering with CRISPR-based counter selection. The tool can be used for generation of multiple mutations in a single day and may open up promising prospective for automation of genome-scale engineering.

Keywords: CRISPR/Cas9, *E. Coli*, Lambda Red, Recombineering, Negative Selection, Web-Tool, Modulation of Protein synthesis

Introduction

E. coli is one of the most widely used model organisms for metabolic engineering for production of a wide range of biochemicals[1]. The integrated use of systems biology, synthetic biology and evolutionary engineering has enabled an extensive portfolio of genetic tools and protocols for efficient, fast and cheap manipulation in order to make *E. coli* suitable for industrial applications. Homologous recombination (RecET)[2], group II intron retrohoming[3–5] and phage-derived recombinases (λ -RED)[6–8] have been applied to introduce modifications in the genome such as single point mutations or Knock-In/Out of genes. However in order to select for the recombinant clones, these methods rely upon antibiotic marker selection, which has to be removed in order to introduce further modifications. Recycling of the selective marker can be achieved by using Cre-Lox recombinase[5] or FLP flippase[4] but both these enzymes leave scars in the genome. This limits the application for sequential allelic exchange since the scars will increase the risk of internal chromosomal rearrangements. In order to overcome the problem of selecting the recombinant strains and recycling of antibiotic markers, different methods of scarless counter-selection have been developed, including a sacB-based method[9] and the use of the meganucleases such as Sce-I[10,11]. Despite the high efficiency of the counter-selection, the repeated engineering process becomes laborious and time consuming with multiple steps involved for the introduction of just a single modification. Additionally, the methods do not enable multiplexing since only few selective markers are available for combinatorial use.

The λ -RED single/double-stranded DNA (ssDNA/dsDNA)-based mediated engineering is currently the only method that allows multiplexing in one step by direct electroporation of single stranded oligos or PCR products. For this reason it has been extensively used as editing tool and it has been employed for Multiplex Automated Genome Engineering (MAGE)[12–14], a method that facilitates genome-scale engineering and barcoded genome profiling. It has recently been improved by co-operative oligonucleotide co-selection by leveraging selectable markers within 500 kb of multiple targets[15]. Moreover it has been demonstrated that it is possible to increase the number of recombinants by using mismatch repair mutants such as *mutS*[16] and by using phosphorothioate linkages to protect the lagging-targeting strand in order to increase the half-life of the Oligos or PCR products and thus the probability of their incorporation during the replication[17]. The efficiency of short ssDNA/dsDNA oligonucleotide-mediated recombineering (including MAGE) is highest for short genome modifications, where around 6 to 20 % can be achieved after single or multiple cycles respectively[13,18], while larger modifications occur with significantly lower frequency (< 1%)[13,17]. Therefore, PCR screening is always required[18] making it challenging to identify the desired mutations that do not result in a clear phenotypic change.

CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats and its associated protein, Cas9) system has recently proven to be a powerful tool for genome engineering in different organism, both Eukaryotes and Prokaryotes such as *E. coli*[19–21] *Actinomycetal*

spp[22]. *Streptomyces spp*.[23,24], lactic acid bacteria[25], *S.cerevisiae*[26–31], higher plants[32,33] *Bombyx mori*[34], Drosophila[35], Zebrafish[36], human cell lines[37,38] and CHO[39–41] cells. The type II CRISPR-Cas system from *Streptococcus pyogenes* consists of the CRISPR- associated (Cas) protein, Cas9, a trans-activating CRISPR RNA (trascrRNA), and a programmable CRISPR targeting RNA (crRNA) that can be fused together to the tracrRNA in a synthetic sgRNA. Cas9 is a programmable nuclease that mediates blunt double-stranded break (DSB) at almost any target DNA locus were a PAM motif (protospacer-adjacent motif) is present at 3' end[42,43]. In the case of *S. pyogenes* nuclease, the sgRNA scaffold can be programmed for a specific site by including 20 bp of the target locus at the 5' position of the double guanine PAM motif (NGG) (20N-NGG). It is also possible to reprogram Cas9 by using tracrRNA and a synthetic array containing 30 bp of the target (5' of NGG) embedded between two repeat regions that will be subsequently be processed in the mature crRNA[42]. In both cases the PAM motif is not included in the target sequence used for the sgRNA or crRNA array. CRISPR/Cas9 can be used to create a selective pressure during the recombineering procedure. The nuclease activity targets the loci that have not incorporated the desired mutation, where it induces double strand breaks (DSB) that have a significant fitness cost and may even result in cell death. In *E. coli*, the CRISPR-Cas9 system has been recently coupled to λ-RED oligos recombineering in order to improve its efficiency[19–21]. Recently, Jiang *et al.*[19] were able to reach high efficiency of recombineering using dsDNA oligos by expressing in succession the two systems. Pyne *et al.* have recently demonstrated that the CRISPR/Cas9 system combined with λ-RED is a promising strategy to facilitate scar-less large chromosomal gene replacement[20] and multiple knock-outs and knock-ins of both small and large fragments[21]. However both of these methods rely mainly on dsDNA oligos, which always require all the λ RED proteins (exo, β and γ) and it is not suitable for a high-through put genome scale approach based on chip-based synthetized oligos[44,45].

Here we aimed at generating a simple and highly effective tool, CRMAGE, which enables fast genome editing. CRMAGE exploits intrinsic negative selection against the wild type of CRISPR/Cas9 in order to increase the MAGE performance for small genome modifications as codon substitution or translation control elements. The system is based on two curable plasmids that encode optimized versions of both systems λ RED recombineering and CRISPR/Cas9. Since both MAGE and gRNA oligos are critical for the protocol, we additionally created a web based tool for automating and optimizing the oligo design. We furthermore demonstrated that degenerate codon usage can be used to expand the range of CRISPR targets, thus creating a powerful tool to modify any site in the genome by introducing different types of modifications such as standard knock-Out/In, point mutations and modulate protein translation in one single step.

Results

Construction of the CRMAGE system

CRMAGE technology exploits the ability of Cas9 to create DSB and thus kill cells that have not had the PAM sequence removed by the recombineering reaction in order to increase the overall MAGE performance. This method has been conceived to compact and optimize at the same time λ RED recombineering and CRISPR/Cas9 in one system ready-to-use and to be applied for automation (Fig. 1). Therefore CRMAGE consist of only two plasmids, which are completely curable (Fig. 1). One plasmid (pMA7CR_2.0) expresses the λ/RED β-protein and the CRISPR/Cas9 protein, which are inducible by L-arabinose and anhydrotetracyline (aTetracycline) respectively. The second vector is a recycling plasmid (pMAZ-SK) that contains an aTetracyline inducible sgRNA used for the negative selection, as well as a "self-destruction" gRNA cassette that targets the vectors own backbone in order to enable plasmid recycling and sequential recombineering steps (Fig. 1). The self-killing system consists on a trascrRNA that combines with two crRNAs, arranged in a natural CRISPR array (30bp repeat – 30bp target - 30bp repeat – 30bp target etc.), to target the pCOLA plasmid origin (Ori) and the kanamycin antibiotic marker upon induction with L-rhamnose and aTetracycline. The λ RED has been coupled to a transient *mutS* phenotype as detailed below, in order to optimize the recombineering step by having a transient deficient repair system. We have also attempted to increase the negative selection of CRISPR/Cas9 by coupling it with a transient *recA*$^{-}$ phenotype in order to promote cell death due to the inability to repair DSB as described below.

Figure 1. Schematic cartoon of CRMAGE system: CRMAGE consist of two plasmids, pMA7CR_2.0 expresses λ/RED β-protein and CRISPR/Cas9 protein that are inducible by L-Arabinose and aTetracyline respectively. The β-proteins are co-expressed with *dam*, which gives a *mutS* mutator phenotype, and *cas9* is expressed in a operon with *recX*, which blocks the repair of double strand breaks. The second plasmid (pMAZ-SK) contains an aTetracyline inducible sgRNA used for selection against the wild type sequence, as well as a self-eliminating circuit that targets its own backbone to enable plasmid recycling and sequential recombineering. Upon L-rhamnose induction (and aTetracycline for *cas9* induction), a trascrRNA that combines with two crRNAs, arranged in a natural CRISPR in order to target the origin (Ori) and the antibiotic marker (Kanamycin).

CRISPR/Cas9 Killing efficiency and estimation of fitness advantage for protocol optimization

In order to investigate the efficacy of the CRISPR/Cas9 negative selection and to estimate the selection time required for the CRMAGE protocol we first determined the CRISPR/Cas9 killing rate and the chance of recombinants take-over using a gRNA targeting the metabolic gene encoding the galactokinase (*galK*). The *galK* gene was used as a target because loss of function of this gene is easy to screen for using McConkey plates, where colonies capable of utilizing for example galactose will turn purple. The CRMAGE system (pMA7CR_2.0 and pMAZ-SK::galK2) was transformed into the WT strain without introducing oligos for recombineering, and we then monitored the number of viable cells over time. An almost complete killing of all cells was achieved only two hours after induction (Fig. 2). Few colonies survived, likely because this CRISPR system may be susceptible to escapers. Indeed only a single point mutation in the PAM motif or in the seed region (8 nucleotides upstream the PAM) can create resistance and thereby eliminating the killing/selection activity. For this reason we tested the optimal time of induction of the CRMAGE system, in order to identify the best window of action where the recombinant population has the most significant fitness advantage over the WT strain. We therefore created a galK mutant that carries a stop codon in position 38 (TAC >TAG) that causes truncation and loss of function of the galactokinase (Tab.1). By inducing the CRMAGE system targeting the wild type *galK* sequence in both a WT background and in a *galK** mutant, it was possible to track the survival of each strain over time. The strain carrying the *galK** mutation took over the population, and more than 90% of the population displayed the mutant phenotype only two hours after induction and no WT colonies were detected after 3 hours.

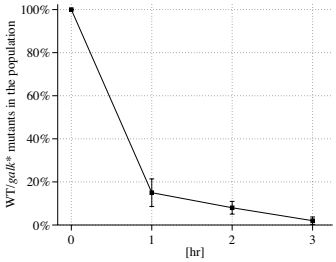


Fig. 2. Wild type killing efficiency: The graph represents the wild type killing efficiency monitored over time. For each time point the cells were plated and the ratio between WT/ *galk** was calculated.

RecA and mutS transient mutants contribution in CRMAGE optimization

In order to increase the efficiency of CRMAGE, we aimed at improving the efficiency of both the λ RED recombineering and the CRISPR/Cas9 killing. Costantino et al. has proven that defects in the mismatch repair system can enhance the efficiency of recombineering ore than 100-fold[16]. We therefore created a transient *mutS*$^{-}$ phenotype by overexpressing the Dam methyltransferase, since this has been documented to result in a mutator (*mutS*$^{-}$) phenotype[46–48]. The *Dam* gene was expressed in an L-arabinose inducible synthetic operon together with the β-protein in order to obtain a transient mutator strain that can be induced only during the recombineering stage, thereby minimizing the creation of unwanted mutations. To optimize the negative selection of the CRISPR/Cas9

system, both *Cas9* and *recX* were expressed in a synthetic operon under control of the pLTet promoter. It has been previously been shown that overexpression of RecX inhibits RecA activity[49,50], one of the major components of the DNA repair system. Since *RecA* mutants are not able to repair DSBs (Double Strand Break), we wished to investigate if overexpression of the inhibitor (RecX) would enhance the CRMAGE negative selection. After one CRMAGE cycle using a galK MAGE oligo we observed that *recX* overexpression most likely has a smaller positive impact on CRMAGE efficiency (p-value <0.1) (Fig. 3), and we therefore decided to keep overexpression of *recX* in the CRMAGE system.

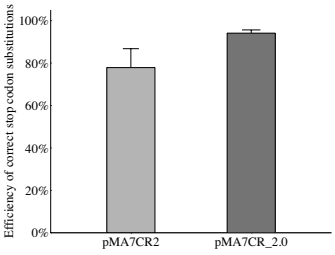


Fig. 3. Contribution of RecX in CRMAGE efficiency: The graph shows the effect of *recX* expression on the CRISPR/Cas9 killing activity after λ-RED recombineering. In pMA7CR_2.0 (represented by the dark grey bar on the right) *cas9* was expressed in a synthetic operon together with *recX*. In pMA7CR2, *cas9* was expressed alone (represented by light grey bar on the left). The presence of *recX* positively contributes to the negative selection (dark grey bar on the right) with p-value < 0.1 according to a T-test analysis.

Use of CRMAGE to target PAM-independent loci

After estimating the optimal killing window for CRMAGE, we wanted to test if the system could be used to target loci in the genome that are not directly linked to a PAM sequence. As proof of concept we decided to create a *galK* knockout by changing one of the initial codons of the gene, which is not directly linked to a PAM site, into a stop codon (TAC >TAG) (Tab.1). To do this, we looked for a PAM in the closest proximity (within the 70 bp coverage of the MAGE oligo) and we introduced a secondary silent mutation that disrupts the nearby PAM sequence while coding for the same amino acid (ACC > ACA) (Tab.1). After a single round of CRMAGE, we found that about 98% (Fig. 4a) of the population had incorporated the desired mutation in *galK*. Using traditional MAGE only 5 % of the population had the specific mutation incorporated (Fig. 4a), which is consistent with previous studies[18] This makes it theoretically possible to utilize the CRMAGE system to target any site in the genome with a very high efficiency thus further expanding the ability to investigate codon usage and re-coding of the essential genes, as it has been previously attempted using MAGE alone[51]

Use of CRMAGE for small insertions and modulation of protein synthesis

The high efficiency of the CRMAGE system for achieving single point mutations prompted us to test if CRMAGE would enable larger genome modifications, which typically result in very low frequencies when relying on MAGE alone[13,17]. Since it is often desirable to modulate expression level of certain proteins, we tested if CRMAGE could be used to modify regulatory elements such as RBS sequences. We initially integrated GFP with a weak RBS on the genome. We then attempted to introduce an RBS variant (TCCTCC >

Locus	Genome Sequence	MAGE OLIGO	gRNA
GFP_RBS CHANGE	//--TGGAGGACTTT--//	//--TGTTCCTCTTT--//	CTCCTTTGCTCATCTAGTAT
Galk2.2_STOP_synonymous codon	//--ACCGACTACA --//	//--ACAGACTAGA--//	AACGAAACCGTCGTTGTAGT

Table 1: Oligos used for the CRMAGE experiments. Taget sites are marked with red. Synonymous mutations are in blue.

AGGAAG) (Tab.1) predicted to have significantly higher expression levels, thus resulting in higher fluorescence levels (GFP^+). Using the CRMAGE system, an efficiency of almost 70% was achieved (Fig. 4b), while only 6% of the population had introduced the RBS modification when relying on MAGE alone (Fig. 4b). This demonstrates that it is feasible to generate libraries of different regulatory elements with high efficiency using CRMAGE

Use of CRMAGE for multiplexing

Given the high efficiency of introducing both single point mutations as well as a larger RBS modification, we decided to test if introduction of multiple mutations could be achieved in a single round of CRMAGE. In a wild type background, we therefore co-selected for the introduction of both the replacement of the RBS sequence in front of GFP and the introduction of a stop-codon in *galK*. The two different MAGE oligos were co-transformed together with two different gRNAs for CRISPR/Cas9 negative selection. The final population was plated on both McConkey and LB plates and the efficiency was determined for each individual mutation (Fig 4c). In randomly picked colonies, 98% were found to carry the *galK** mutation and 70% carried the RBS modification (Fig 4c). Subsequently, recombinants clones for each mutation (GFP^+ and *galk**) were tested if they also carried the second mutation. All the tested colonies (40 out of 40) showing strong GFP fluorescence after CRMAGE also resulted in white colonies when plated on McConkey plates, thus proving that they also incorporated the *galK** mutation. Similarly, all the colonies tested (15 out of 15) carrying the galK* mutations also showed strong GFP expression, thus resulting in a 100% efficiency of simultaneously introducing the double mutations (Fig. 4d). Using MAGE alone, only 5-6% of the colonies carried the galK* mutation, and of these only 10% carried showed strong GFP fluorescence (Fig. 4d).

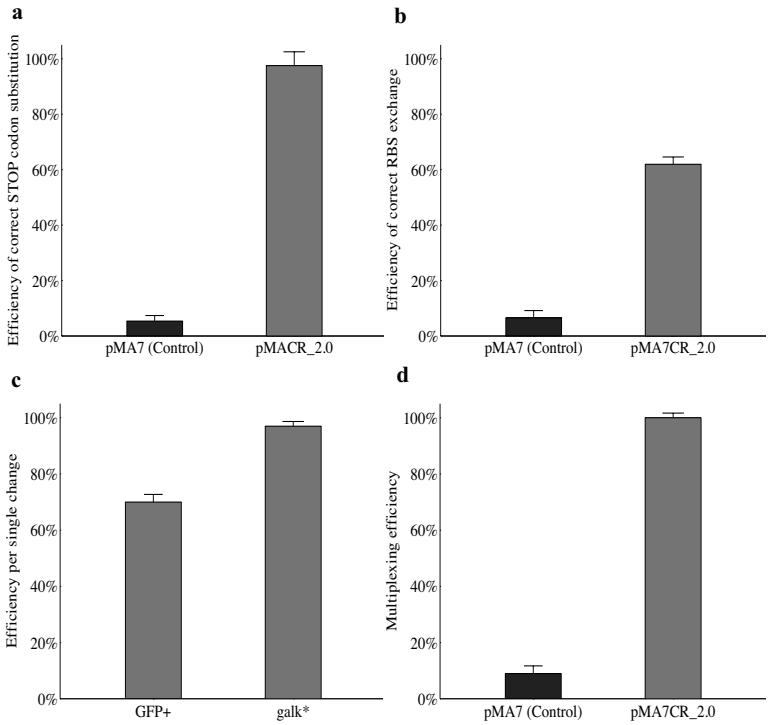


Figure 4: Efficiency of CRMAGE. Panel A and B show the efficiency of CRMAGE for gene recoding using pMA7CR_2.0 compared to the pMA7 control using only λ-RED. (**a**) The efficiency of CRMAGE for introducing codon substitution (ga*lk** TAC >TAG) linked to a secondary silent codon substitution used by CRMAGE as counter selection (ACC > ACA). (**b**) The efficiency of substituting a weak RBS in front of GFP with a strong one (TCCTCC > AGGAAG) while disrupting the PAM sequence used for negative selection. Panel C-D shows the efficiency of multiplex CRMAGE. (**c**) The efficiency of a single round of CRMAGE for simultaneously introducing the two mutations mentioned above (the RBS exchange on the left and stop codon substitution on the right). (**d**) CRMAGE multiplexing efficiency (pMA7CR_2.0) of the two mutations compared to the control (pMA7) using only λ-RED. The data result from the analysis of 15 to 40 positive clones per each mutation that were re-streaked and tested if they were carrying also the second modification.

Elimination of gRNA plasmid for subsequent rounds of CRMAGE

In order to enable sequential recombineering cycles, it is beneficial if the plasmid required for the negative selection can be quickly cured, so that a new vector can be introduced for the next round of CRMAGE. For this reason we designed the plasmid carrying the gRNAs as a self-killing plasmid (pMAZ-SK) by exploiting CRISPR/Cas9 ability to cut the plasmid DNA. An L-rhamnose inducible CRISPR natural array encoding two pre-crRNAs that target the origin of the plasmid and the kanamycin antibiotic marker was therefore included in the vector. In order to create mature active crRNA, the trascrRNA was included on the plasmid under the control of a strong synthetic constitutive promoter. This way it should possible to induce plasmid digestion in order to cure the plasmid before starting a new round of CRMAGE. We initially determined the minimum amount of time required for successful curing of the plasmid by following the number of cells that retained the plasmid after L-rhamnose induction (Fig. 5). After just 2-3 hours of induction, 92-96% of the cells had lost the plasmid. Interestingly, a complete removal of the plasmid could not be achieved even after overnight incubation, where 0.2% of the population still retained the plasmid (Fig. 5). However, for the practical use for CRMAGE, very efficient plasmid loss can be achieved even after a short incubation time, which will significantly speed up the engineering process. The effective plasmid loss induced by the expression of trascRNA and the CRISPR array that targets the plasmid origin and the kanamycin cassette demonstrates that the processing of the RNA is functional. Using the same principle, it will therefore also be possible to include gRNAs that target the pMA7CR_2.0 plasmid, thereby making it possible to remove both plasmids at the end of the final engineering cycle.

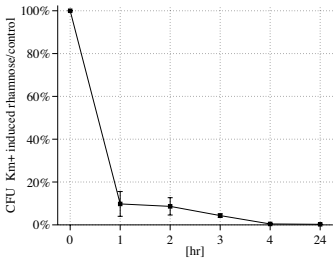


Figure 5: Self-eliminating proprieties of pMAZ-SK plasmid used for negative selection. The graph shows the rate of plasmid curing over time. The population was induced with L-rhamnose (to induce self-killing crRNA) and aTetracycline (to induce Cas9 expression) and was plated on Kanamycin selective plates and compared to the un-induced population.

CRMAGE gRNA design tool

A number of tools are available online for identifying and designing gRNAs for various organisms. However, since CRMAGE enables the generation of specific mutations throughout the genome, it is necessary to both identify the optimal gRNA for counter selection, as well as the oligos required for MAGE. Furthermore, when the specific mutation is not located in close proximity of a PAM sequence, it is necessary to identify a nearby PAM sequence that can be included in the MAGE oligo design as described earlier. In order to automate the design of both sets of oligos, we have created a web based tool, CRMAGE – Web tool (available at http://staff.biosustain.dtu.dk/laeb/crmage/), which guides the user through the necessary steps to create a CRMAGE mutated oligos, considering also the option of using

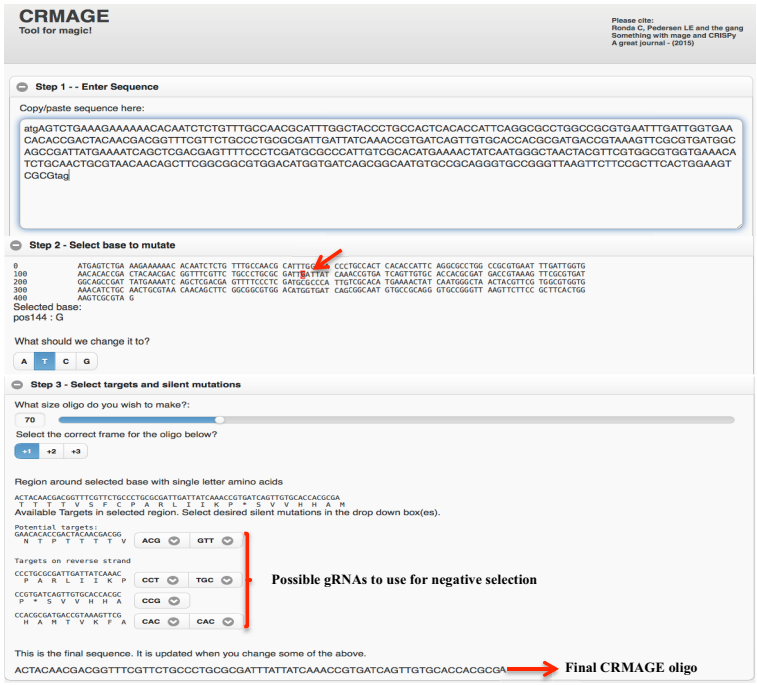


Figure 6: CRMAGE Web-Tool. The program guides the user through the necessary steps to create a CRMAGE mutated oligos, considering also the option of using a degenerate codon if the desired mutation does not contain a PAM itself. It consists of three major steps: (1) Enter (copy/paste) the wild type DNA sequence; (2) Select the base to mutate and which base to change to. Step 3: Choose the Open Reading Frame and the size of the CRMAGE mutation oligo. A list of potential gRNA sites are shown, and the user can pick silent mutations to destroy the indicated gRNA PAM site (pointed by the red bracket). Finally the program outputs the CRMAGE oligo sequence (shown by the red arrow).

degenerate codons if the desired mutation is not located near a PAM sequence. It consists of 3 major steps which include: input of a DNA sequence to mutate, user based selection of the base(s) to mutate, and the program finally presents the user with available gRNAs near the desired mutation spot as well as with options to mutate the PAM site of the gRNA to prevent it from selecting against the recombinants. The final output includes the gRNA as well as an SS-DNA oligo to use for recombineering that contains both the desired mutation and the one required for negative selection (Fig. 6).

DISCUSSION

CRMAGE is based on the combination of MAGE recombineering with the negative selection against the wild type sequence, which can be achieved using CRISPR/Cas9 targeted double strand breaks. In a single round of CRMAGE recombineering, we were able to achieve close to 98% efficiency for single point mutations versus 5% using traditional MAGE. For replacing a larger 6 bp RBS sequence, an average of 66% was achieved in two separate types of experiments, while only 6% efficiency was reached using traditional MAGE. Such high efficiency with ssDNA oligos and short homology arms (70 bp) has not been previously reported. Pyne *at al*.[20] achieved only 39% for small insertions (8bp) using dsDNA and 80% for small modifications with ssDNA and Jiang *et al*.[21], could reach efficiencies in the range of 72% to 92% relying mainly on long homology arms (300-400pb), while focusing mainly on large modifications. Moreover CRMAGE system enables integration of multiple mutations with no apparent loss of efficiency, which is different from previous work where the multiplexing efficiency dropped considerably from 39-47% to 0.68-0.25% and from 72-92% to 48-78% in Pyne *et al.* and Jiang et all systems respectively[20,21]. In the case of CRMAGE, we speculate that the reason for incomplete killing of the WT or rescue of the selection plasmid (pMAZ-SK) is the leakiness of the tetracycline and rhamnose promoters that may result in a selective pressure during the 5 hours of the whole CRMAGE protocol, which may result in few escapers that will not succumb to the negative selection.

It has previously been attempted to modulate chromosomal gene expression in one single step using PCR products[52]. However this work relied on phenotypic screening, and the method leaves FRT scars after the regulatory element is inserted. The presence of scars in the genome promotes genome rearrangement after several rounds of modifications thus making it difficult to further engineer the strain. With CRMAGE we have demonstrated that is possible to modulate the protein translation with relatively high efficiency within the population, which can be applied to any regulatory element that could be inserted to modulate either the gene expression or protein translation.

The CRMAGE system was furthermore designed to enable efficient recycling of the plasmid used for negative selection by targeting the Cas9 towards the origin and kanamycin cassette in the vector backbone itself. Controlled by a L-rhamnose inducible promoter, an almost complete loss (96%) of the plasmid was achieved after 2-3 hours of induction. This means that it is possible to proceed with subsequent rounds of CRMAGE without an intermediate screening for kanamycin sensitive clones. Using the same approach it will also be possible to remove the entire CRMAGE system in the last CRMAGE round, where the CRISPR/Cas9 killing activity can target origin and antibiotics markers in both plasmids thus resulting in a clean recombinant strain. The idea of generating a clean background strain and the possibility to recycle the plasmid containing the gRNA at the end of the engineering process has been considered by Jiang *et al.*[21]. In their work they have aimed at rescuing the plasmid harboring the gRNA for negative selection by induction with IPTG, and the

plasmid carrying λ RED and Cas9 by growing the culture at 37℃. However it was not possible to proceed with a subsequent round of recombineering before of 8 or 16 hours of IPTG induction. The use of the temperature sensitive origin used for plasmid encoding Cas9 furthermore increases the time required for one single round of recombineering since the growth rate of *E. coli* is significantly reduced at 30℃.

For multiplexing of the CRMAGE targets, it is necessary to express multiple synthetic gRNAs that have significant stretches of sequence homology (typically 136 nucleotides). As discussed by Jiang *et al.*, the repeated sgRNAs may result in homologous recombination[21]. Here, we have shown that the presence of a constitutively transcribed trascRNA in the pMAZ-SK vector makes it possible to generate multiple gRNAs from a single short region carrying only 30 nucleotides per gRNA followed by a 30 nucleotide repeat between them. The synthesis and cloning of such an array of gRNAs (target-repeat-target-repeat etc.) is simple and it minimizes the risk of recombination in the vector. The expression of the pre-cRNA for multiplex CRMAGE may substitute the sgRNAs if put under control of the inducible pLtet promoter, making it easy to control the timing of the expression. The system will most optimally use aTetracycline induction for negative selection (when pLtet is used in front of the CRISPR array) and L-Rhamnose induction for curing the plasmid.

We furthermore demonstrate that degenerate codon usage can be used to expand the range of CRISPR targets, thus creating a powerful tool for targeting virtually any site in the genome and enabling single step engineering to create single point mutations as well as larger mutations, as demonstrated by the efficient RBS replacement for modulating protein translation. The same design can be applied for creating knock-ins or complete knock-outs, depending on the donor oligos. The web based CRMAGE web - tool has been engineered to facilitate the design of both the λ red oligo required for the specific mutation as well as gRNA required for the negative selection and together with the convenient USER-cloning-assembly-based system we have developed, speed up the design and cloning process. Since CRMAGE enables the generation of multiple mutations in a single cycle and multiple cycles within a working day, it has the potential to significantly increase the daily strain engineering capacity. The increased efficiency furthermore opens up the possibility of automating genome-scale engineering.

MATERIALS AND METHODS

Strains, Media and Reagents

E.coli K-12, MG1655 strain with genome integrated repressor from pZS4Int-tetR, was used to perform CRMAGE experiments and Dh5α strain instead was used for cloning purposes. CRMAGE was performed using LB-Lennox (10 g/L tryptone (“Enzymatic digest from caseine”- Fluka Analytical), 5 g/L yeast extract (Fluka Analytical), 5 g/L NaCl) supplemented with 0.5 mM of $MgSO_4$ (Sigma). All cultures were grown at 37℃, 250rpm shaking. For cloning Q5 high fidelity Polymerase (NebLab), Fasta Digestion enzymes and Buffer from Fermantas/Thermo Scientific and USER enzymes from NebLab was used. All oligonucleotides and gblocks

were synthesized by Integrated DNA Technologies (IDT; Coralville, IA) at the 25 nm scale using standard desalting.

Plasmids constructions and CRMAGE plug-in and play BioBrick

The pMA7CR_2.0 (Amp^R) has been constructed by cloning Cas9 and RecX in pMA7 plasmid [53]. *RecX* was first amplified from purified MG1655 genome using the primers 1 and 2 (Table S1) and cloned in pZ21MCS[50] downstream of pLTet promoter using the restriction enzymes KpnI and BamHI. Then the expression cassette, gene + promoter (pLtet_recX), was amplified and cloned in pMA7[53] with USER using primers 3, 4 and 5, 6 (Table S1) for the pMA7 backbone and RecX respectively. Eventually a synthetic operon was created by cloning Cas9 with a strong RBS (AAGGAGA) downstream pLTet_recX using primers 7, 8 (Table S1) to amplify the backbone with recX, and 9, 10 (Table S1) to amplify Cas9 directly from *S.pyogen* genome (ATCC collection) (Fig. S1). The plasmid pMACR_2.0 comes in combo with plasmid pZS4Int-tetR (CAM^R)[54] that is a single copy plasmid harboring TetR repressor necessary for the control of Cas9 and sgRNA expression. This plasmid can also be integrated in the genome in one single step using the helper plasmid pLDR8 [WM2269] (ATCC® 77357) expressing the integrase. The self-killing gRNA plasmid (pMAZ-Self-Killing: pMAZ-SK) was designed using pCOLA-duet (Millipore) as backbone vector and it consists on two main parts. The first one contains the aTetracycline inducible synthetic gRNA use for negative selection and the second one constituted of two subparts for the self-killing (Fig. 1), the pre-crRNA array and the trascrRNA. The pre-crRNA array and the trascrRNA were ordered as synthetic oligos and gBlock from Life Technology and IDT respectively. The first one, named part1_selfkilling (Table S1) was designed to have the pre-crRNA array coding for two crRNA targeting the Origin (Ori) and the antibiotic marker (Kanamycin) of the pCOLA backbone, and it was cloned using the restriction enzymes *Nco*I and *Pac*I after the part containing the trascrRNA (part2_selfkilling) was cloned with USER using primers 11 to 14 (Table S1). In the part2_selfkilling component, there is the trascrRNA under a constitutive synthetic promoter and a synthetic terminator, both derived from igem.part.org (Table S1). Once the two major components of the self-killing plasmid were assembled in the pCOLA backbone, the synthetic gRNA for the negative selection was added using primers 17 to 20 (Table S1) with USER cloning[55]. Only the gRNA GFP4 (Table S1) was ordered as a gBlock (from IDT) to have it as a unit organized in: the aTetracycline inducible promoter (pLtet), 20pb target sequence and the general scaffold that fuses part of the repeat region of the crRNA and the trascrRNA.

In order to change the target sequence in the synthetic gRNA we have designed a USER-cloning-assembly-based plug-and-play system, where it is only necessary to order two oligos to change the synthetic gRNA for the negative selection. This system relies on USER cloning[55]: the self-killing backbone was amplified using the universal primers 21, 22 (Table S1) containing the uracil, the target sequence can be replaced by changing only 20 internal nucleotides of two universal complementary oligo scaffolds (oligo

scaffolds 23, 24 shown in Table S1). These oligos were designed to be complementary and when annealed will leave 10pb overhangs that match the overhangs left on the backbone after USER treatment. Therefore it is necessary replace only the 20 bp sequence with the new target in a way that the two oligos sequences will be complementary (20pb sequence on the forward oligo has to be complementary to the 20bp sequence on the reverse oligo). The oligos were ordered without uracil making the method very cost efficient, and 10 µL of each oligo (100µM concentrated) was then mixed with 10 µL of Neb Buffer4 and 70 µL of MilliQ Water (100 µL reaction in total). The reaction mix was incubated at 95 °C for 5 minutes and let it cool down slowly overnight. Once the two oligos were annealed, they were mixed with the amplified backbone, treated with USER enzymes and directly transformed into chemically competent cells. The self-killing plasmid containing the galk2.2 gRNA was obtained with this method using primers 24, 25 (Table S1) in just a single day of work. The second gRNA for multiplexing was inserted in pMAZ-SK::galK2 using primers 27 to 30 (Table S1). pMA7CR_2.0 and pMAZ-SK sequences are shown in Table S1.

CRMAGE protocol

An overnight culture of MG1655::pMA7RCR_2.0 in LB with 100 µg/mL ampicillin (to maintain pMA7CR_2.0), 35 µg/mL chloramphenicol (to maintain pZS4Int-tetR if not integrated in the genome) was shaken at 37°C. The following day 15 mL of LB with 100 µg/mL ampicillin, 35 µg/mL chloramphenicol was inoculated with 0.15 mL of the overnight culture in a 250 mL flask (starting OD between 0.02-0.03), incubated at 37°C shaking and grown until an OD of 0.5-0.6 (about 1.5-2 hours). λ -RED was induced by adding L-arabinose to a final concentration of 0.2%, followed by continued shaking at 37°C for 10-15 minutes. The culture was immediately put in a ice-water bath and left to cool for at least 15-20 minutes. The culture was then pelleted by centrifugation at 6500 x *g* for 5 minutes at 4°C, the supernatant was discarded and the pellet resuspended in 35 mL of ice-cold MilliQ water. This procedure was repeated two times more, where the pellet was first re-suspended in 15 mL and then in 1 mL of ice-cold MilliQ water. The 1 mL of cells were finally centrifuged at 6500-7000 x *g* for 5 minutes at 4°C and the supernatant removed completely. Subsequently, cells were prepared for electroporation and re-suspended in 400-800 µL of ice-cold MilliQ water. The MAGE oligonucleotides together with the plasmid for negative selection were prepared in advance in an Eppendorf tube containing 0.5 µL of equimolar amounts of each oligo (100 pmol/µL of each) and 250 ng of the plasmid in MilliQ water. 50 µL of cells were added to the oligos/plasmid mix and electroporated at 1.8 kV in a 1 mm gap cuvette, and 950 µL of LB with 100 µg/mL ampicillin, 35 µg/mL chloramphenicol was added immediately after and the cells were transferred to a new tube (15 mL) and left to recover for 1h at 37°C with shaking. Kanamycin was added to reach a concentration of 50 µg/mL and the culture was incubated for additional 2 hours.

After 3 hours of incubation, tetracycline (200ng/mL) was added and the cells were grown for another 2 hours at 37°C shaking. After this point the cells can either be plated on selective media or the plasmid for the

negative selection can be rescued to start another round of MAGE. For the plasmid cure, the cells were washed twice with fresh LB and resuspended in LB with 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, aTetracycline 200ng/mL and 0.2% (w/v) of L-rhamnose. If the cell were left to grow ON, a smaller inoculum (10^3 - 10^4 dilution) was used. If the culture was prepared for an immediate subsequent round of CRMAGE, the it was diluted to an OD that allows at least 2-3h of growth before reaching OD:0.5-0.6, which is necessary to start the following round.

CRMAGE Web-Tool

The CRMAGE web-based design tool is available at http://staff.biosustain.dtu.dk/laeb/crmage/ and follows three individual steps as described here: Step 1: Enter the wild type DNA sequence. Simply copy/paste or manually enter the sequence. Step 2: Select the base to mutate. Here the user is presented with a numbered and spaced version of the sequence entered in step 1 and the user can then click the exact base that should be mutated and can select into which base it should be mutated. Step 3: CRMAGE mutation oligo and gRNA. In this step the user can select a desired CRMAGE mutation oligo size. The size of the MAGE oligo determines which gRNA sites can be used since the program sets that as constraint region from which it samples the possible gRNAs to use as "negative selection marker". The user should then select the correct reading frame. Then a list of potential gRNA sites are shown and the user can pick silent mutations with the intent to destroy the indicated gRNA PAM site. Be aware that all silent mutations are shown even if they do not destroy the gRNA PAM site. It is up to the user to ensure that the PAM site is destroyed. One or two mutation boxes may be shown next to a gRNA target. This is dependent on whether one or two codons cover the PAM site. At the very end of step 3, the mutation oligo sequence is presented and can be copy/pasted for synthesis.

Competing interests

The authors declare no competing interests.

Authors' contributions

CR and ATN conceived the study; CR performed all the experiments and analyzed the data; CR generated plasmids; CR conceived the Web-Tool and LEP wrote the scripts and made the web interface. CR drafted the manuscript and ATN contributed to prepare the final version of the manuscript.

Acknowledgements

This work was financed by The Novo Nordisk Foundation. We thank Prof. Morten Otto Alexander Sommer, Mads Bonde, Dr. Rebecca Lennen, for scientific discussion. Dr. Margite Pedersen and Dr. Annika Nilsson Wallin are thanked for providing the pMA7 plasmid containing L-rhamnose inducible λ-RED and Dam.

Bibliography

1. Chen, X. *et al.* Metabolic engineering of Escherichia coli: a sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* **31,** 1200–1223 (2013).

2. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20,** 123–128 (1998).

3. Karberg, M. *et al.* Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.* **19,** 1162–1167 (2001).

4. Sukhija, K. *et al.* Developing an extended genomic engineering approach based on recombineering to knock-in heterologous genes to Escherichia coli genome. *Mol. Biotechnol.* **51,** 109–118 (2012).

5. Enyeart, P. J. *et al.* Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol. Syst. Biol.* **9,** 685 (2013).

6. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–6645 (2000).

7. Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci.* **97,** 5978–5983 (2000).

8. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4,** 206–223 (2009).

9. Yu, B. J. *et al.* Rapid and efficient construction of markerless deletions in the Escherichia coli genome. *Nucleic Acids Res.* **36,** e84–e84 (2008).

10. Posfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F. R. Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* **27,** 4409–4415 (1999).

11. Yang, J. *et al.* High-efficiency scarless genetic modification in Escherichia coli by using lambda red recombination and I-SceI cleavage. *Appl. Environ. Microbiol.* **80,** 3826–3834 (2014).

12. Wang, H. H. & Church, G. M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol.* **498,** 409–426 (2011).

13. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).

14. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28,** 856–862 (2010).

15. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* **40,** e132–e132 (2012).

16. Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci.* **100,** 15748–15753 (2003).

17. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186,** 791–799 (2010).

18. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci.* **98,** 6742–6746 (2001).

19. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

20. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system to lambda Red recombineering enables simplified chromosomal gene replacement in Escherichia coli. *Appl. Environ. Microbiol.* (2015).

21. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).

22. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* (2015).

23. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2015).

24. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 231–243 (2015).

25. Oh, J. H. & van Pijkeren, J. P. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic Acids Res.* **42,** e131–e131 (2014).

26. DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41,** 4336–4343 (2013).

27. Bao, Z. *et al.* Homology-Integrated CRISPR{\textendash}Cas (HI-CRISPR) System for One-Step Multigene Disruption in Saccharomyces cerevisiae. *ACS Synth. Biol.* (2014). at <doi/abs/10.1021/sb500255k>

28. Horwitz, A. A., Walter, J. M., Schubert, M. G. & Kung, S. H. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst.* (2015).

29. Jakočiunas, T. *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metab. Eng.* (2015).

30. Stovicek, V., Borodina, I. & Forster, J. CRISPR-Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains. *Metab. Eng. Commun.* **2,** 13–22 (2015).

31. Ronda, C. *et al.* CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. *Microb. Cell Fact.* **14,** 97 (2015).

32. Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 686–688 (2013).

33. Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* **41,** e188–e188 (2013).

34. Wang, Y. *et al.* The CRISPR/Cas system mediates efficient genome engineering in Bombyx mori. *Cell Res.* **23,** 1414–1416 (2013).

35. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep.* **4,** 220–228 (2013).

36. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res* **23,** 465–472 (2013).

37. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).* **339,** 819–823 (2013).

38. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

39. Ronda, C. *et al.* Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* **111,** 1604–1616 (2014).

40. Lee, J. S., Kallehauge, T. B., Pedersen, L. E. & Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* **5,** 8572 (2015).

41. Grav, L. M. *et al.* One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* (2015).

42. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

43. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337,** 816–821 (2012).

44. Cobb, R. E., Si, T. & Zhao, H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr. Opin. Chem. Biol.* (2012). doi:10.1016/j.cbpa.2012.05.186

45. Esvelt, K. M. & Wang, H. H. Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* **9,** 641 (2013).

46. Herman, G. E. & Modrich, P. Escherichia coli K-12 clones that overproduce dam methylase are hypermutable. *J. Bacteriol.* **145,** 644–646 (1981).

47. Yang, H., Wolff, E., Kim, M., Diep, A. & Miller, J. H. Identification of mutator genes and mutational pathways in Escherichia coli using a multicopy cloning approach. *Mol. Microbiol.* **53,** 283–295 (2004).

48. Robbins-Manke, J. L., Zdraveski, Z. Z., Marinus, M. & Essigmann, J. M. Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient Escherichia coli. *J. Bacteriol.* **187,** 7027–7037 (2005).

49. Stohl, E. A. *et al.* Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo. *J. Biol. Chem.* **278,** 2278–2285 (2003).

50. Drees, J. C., Lusetti, S. L. & Cox, M. M. Inhibition of RecA protein by the Escherichia coli RecX protein: modulation by the RecA C terminus and filament functional state. *J. Biol. Chem.* **279,** 52991–52997 (2004).

51. Lajoie, M. J. *et al.* Probing the Limits of Genetic Recoding in Essential Genes. *Science (80-.).* **342,** 361–363 (2013).

52. Meynial-Salles, I., Cervin, M. A. & Soucaille, P. New tool for metabolic pathway engineering in Escherichia coli: one-step method to modulate expression of chromosomal genes. *Appl. Environ. Microbiol.* **71,** 2140–2144 (2005).

53. Sandberg, T. & Pedersen, M. Evolution of Escherichia coli to 42° C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. ... *Biol. Evol.* 1–38 (2014). doi:10.1093/molbev/msu209

54. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).

55. Nour-Eldin, H. H., Geu-Flores, F. & Halkier, B. A. USER cloning and USER fusion: The ideal cloning techniques for small and big laboratories. *Methods Mol. Biol.* **643,** 185–200 (2010).

Supplementary Figures, Tables and Sequences

Table. 1

	Primer		Source
1	recX_fwd_kpnI	ATCGAGGTACCATGACAGAATCAACATCCCGTC	This study
2	recX_rev_BamHI	TCGTAGGATCCATCAGTCGGCAAAATTTCGCCA	This study
3	pMA7_fwd	AAGCGGGUTTTTTTATGACAAACTCTTTTGTTTATT	This study
4	pMA7_rev	ATAGGGAATAGGGAGUAGAAACGCAA	This study
5	fwd_insert_recX	ACTCCCTATTCCCTAUCAGTGATAGAGATTGACAT	This study
6	rev_insert_recX	ACCCGCTUGCGCGGGCTTTTTCACATTGATGCCTCTAG CACGC	This study
7	pMA7CR_2.0_fwd	ACTGAGA/ideoxyU/CCCATGGTACGCGTG	This study
8	pMA7CR_2.0_rev	ATCTCCTTC/ideoxyU/caGTCGGCAAAATTTCG	This study
9	c9_PMA7CR_2.0 fwd	AGAAGGAGA/ideoxyU/ATACATGGATAAGAAATACT	This study
10	c9_PMA7CR_2.0 rev	ATCTCAG/ideoxyU /CACCTCCTAGCTGACTCA	This study
11	trancrRNA_Coli_fwd	ACACCGAC/ideoxyU/AGCGAAAAAACC	This study
12	trancrRNA_Coli_rev	ACGCTGCT/ideoxyU/TTGACGGCTA	This study
13	pCOLAtran_rev	AGTCGGTG/ideoxyU/GCGCAACGCAATTAATGTA	This study
14	pCOLAtran_fwd	AAGCAGCG/IDEOXYU/ATATACCATGGGCAGCAGC	This study
15	Part2_selfkilling	TTAATTAACACCGACTAGCGAAAAAACCCCGCCGAAG CGGGGTTTTTTGCGAAAAAAAGCACCGACTCGGTGCCA CTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT GCTATGCTGTTTTGAATGGTTCCGCTAGCACTGTACCTA GGACTGAGCTAGCCGTCAAAAGCAGCGTCCTCAGG	This study
16	Part1_selfkilling	CCATGGCCACAATTCAGCAAATTGTGAACATCATCACG	This study

		TTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTC AGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGAC TGGTCGTGTTTTAGAGCTATGCTGTTTTGAATGGTCCCA AAACCCGTTCCGTGTAGACAGTTCGCTCCAAGCTGTTT TAGAGCTATGCTGTTTTGAATGGTCCCAAAACGCCATG TTTCAGAAACAACTCTGGCGCATCGTTTTAGAGCTATG CTGTTTTGAATGGTCCCAAAACAAAAAAAAACCCCGCC CCTGACAGGGCGGGGTTTTTTTTTTAATTAA	
17	pCOLAforgrna_fwd	ACCGGTA/ideoxyU/CCTAGGCTGCTGCCACCG	This study
18	pCOLAforgrna_rev	ACCAGAC/IDEOXY U/TTAATTAAAAAAAAAACCCCGCCCTGTCA	This study
19	pCOLA_gRNA_fwd	AGTCTGG/ideoxyU/TATAACCTGAGGTCCCTATCAGTGA TAGAGA	This study
20	pCOLA_gRNA_rev	ATACCGG/ideoxyU/TTCGACTTAAGCATTATGCGG	This study
21	new universal backbone fwd primer	AGCTAGAAA/ideoxyU/AGCAAGTTAAAATAAGGC	This study
22	New rev back bone primer	AGTATCTC/ideoxyU/ATCACTGATAGGGATGTCA	This study
23	variable region fwd	GAGCAC (20N) GTTTTAGAGCTAGAAAT	This study
24	Variable region rev	CTAAAAC(20N)GTGCTCAGTATCTCT	This study
25	Variable region galk2 fwd	**GAGCAC**AACGAAACCGTCGTTGTAGT**GTTTTAGAGCT AGAAAT**	This study
26	Variable region galk2 rev	**CTAAAAC**ACTACAACGACGGTTTCGTT**GTGCTCAGTA TCTCT**	This study

27	2_sgRNA_fwd	**ACCGGTATTCCCTATCAGTGATAGAGAT**	This study
28	2_sgRNA_rv	**AGCAGCCTAGGAAAAAAGCACCGACTCGG**	This Study
29	2_sgRNAbb_fwd	**AGGCTGCTGCCACCGCTGA**	This Study
30	2_sgRNAbb_rev	**ATACCGGTAAAAAAGCACCGACTCGGTGCCA**	This Study
	CRMAGE oligos		
	GFP_RBS CHANGE	GACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTCATCTA GTATTGTTCCTCTTTAATCTCTAGTAGCTAGCACTGTAC CTAGGACTGAGC	This study
	Galk2.2_STOP_synonymous codon	GGCCGCGTGAATTTGATTGGTGAACACACAGACTAGA ACGACGGTTTCGTTCTGCCCTGCGCGATTGATT	This study
	Synthetic gRNA		
	Promoter (pLtet)	TCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGA TAGAGATACTGAGCAC	This study [54]
	Scaffold	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTTTT	[37,38]
	GFP4 target	CTCCTTTGCTCATCTAGTAT	This study
	Galk 2 target	AACGAAACCGTCGTTGTAGT	This study
	crRNA Array		
	pRham	CCACAATTCAGCAAATTGTGAACATCATCACGTTCATC TTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAAC GAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCG T	This study
	Repeat	GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC	This study
	Target Ori	CCGTTCCGTGTAGACAGTTCGCTCCAAGCT	This study
	Target Kanamycin	GCCATGTTTCAGAAACAACTCTGGCGCATC	This study
	Synthetic Terminator for crRNA array(BBa_b1006)	AAAAAAAAACCCCGCCCCTGACAGGGCGGGGTTTTTTT T	This study

	trascrRNA		
	trascrRNA sequence	GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG GTGCTTTTTTT	This study,[42]
	Promoter for trascrRNA (BBa_J23100pr)	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC	This study
	Terminator for trascrRNA (BBab_1002)	CGCAAAAAACCCCGCTTCGGCGGGGTTTTTTCGC	This study

pMA7CR_2.0 sequence:

CCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAA

ACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGA

CCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGG

TAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGC

ATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTC

GTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGAC

AAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCC

CGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGC

ATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTG

CGTTTCTACTCCCTATTCCCTAUCAGTGATAGAGATTGACATC

CCTATCAGTGATAGAGATACTGAGCACATCAGCAGGACGCAC

TGACCGAATTCATTAAAGAGGAGAAAGGTACCatgACAGAATC

AACATCCCGTCGCCCGGCATATGCTCGCCTGTTGGATCGTGC

GGTACGCATTCTGGCGGTGCGCGATCACAGTGAGCAAGAACT

GCGACGTAAACTCGCGGCACCGATTATGGGCAAAAATGGCCC

AGAAGAGATTGATGCTACGGCAGAAGATTACGAGCGCGTTAT

TGCCTGGTGCCATGAACATGGCTATCTCGATGACAGCCGATT

TGTTGCGCGCTTTATCGCCAGCCGTAGCCGCAAAGGTTATGG

ACCTGCGCGTATTCGCCAGGAACTGAATCAGAAAGGTATTTC

CCGCGAAGCGACAGAAAAAGCGATGCGTGAATGTGACATCGA

CTGGTGCGCACTGGCGCGCGATCAGGCGACGCGAAAATATG

GCGAACCTTTGCCAACTGTCTTTTCAGAAAAAGTTAAGATCCA

GCGTTTTCTGCTCTATCGTGGCTATCTGATGGAAGATATCCAG

GAGATTTGGCGAAATTTTGCCGACtgAGAAGGAGATATACATG

GATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGC

GTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCT

AAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATC

AAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAG

ACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGG

TATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTT

TTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCG

ACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGA

ACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTAT

CATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGG

TAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGC

CTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAG

GGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTA

TCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCC

TATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGC

ACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCA

GCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCAT

TGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTT

GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTT

ACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCA

ATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCT

ATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTA

AGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAAC

ATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACA

ACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAA

AACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAA

GAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATG

GTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGC

TGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATC

AAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAG

AAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGA

AAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTG

GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCT

GAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGAT

AAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACT

TTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAG

TTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAG

GTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTT

CAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAA

CAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTT

CAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTT

GAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGC

TAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAA

TGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTT

GAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCT

CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGC

CGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATG

GTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT

GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATC

CATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCAC

AAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAA

ATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGAC

TGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCA

TAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCA

GACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAA

ACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCT

TAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAA

GCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTG

GACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCG

ATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGA

CAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATC

GGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAA

CTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGT

AAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGT

GAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAA

CTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTC

GCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA

GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTC

CGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATT

ACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAA

CTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGT

CTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCT

AAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTC

TTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACAC

TTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTA

ATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATT

TTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATA

TTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGG

AGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCG

TAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAG

TCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGA

AAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACT

AGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCC

GATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAA

AGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTA

GAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTA

CAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAAT

TTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTC

CAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATA

AGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTC

TAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTT

AGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAA

GCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAG

CTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAA

ACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATC

CATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGA

GTCAGCTAGGAGGTGACTGAGATCCCATGGTACGCGTGCTAG

AGGCATCAATGTGAAAAAGCCCGCGCAAGCGGGUTTTTTTAT

GACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTA

TCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATAT

TGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCC

CTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCA

CCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTT

GGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGG

TAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATG

ATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCC

GTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACT

ATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA

AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTG

CTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT

GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCA

CAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACC

GGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAC

GATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAAC

TGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC

TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG

GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCC

GGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCA

GATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGG

AGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG

ATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA

GTTTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAGC

GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT

GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC

TTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGT

CAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGT

GCTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGAT

GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGC

CCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT

TCCAAACTTGAACAACACTCAACCCTATCTCGGGCTATTCTTT

TGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAA

AATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAA

TATTAACGTTTACAATTTAAAAGGATCTAGGTGAAGATCCTTT

TTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTT

CCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC

TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACA

AAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAA

GAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGA

GCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTA

GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC

GCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT

AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCG

GATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCAC

ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA

CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG

GAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAA

CAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGG

TATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGC

GTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGA

AAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT

GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGA

TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC

GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG

CGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC

GCATCTGTGCGGTATTTCACACCGCATAGGGTCATGGCTGCG

CCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCT

TGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTC

TCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG

AAACGCGCGAGGCAGCAAGGAGATGGCGCCCAACAGTCCCC

CGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGC

TCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGA

TGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGG

TGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCTGCTCAT

GTTTGACAGCTTATCATCGATGCATAATGTGCCTGTCAAATGG

ACGAAGCAGGGATTCTGCAAACCCTATGCTACTCCGTCAAGC

CGTCAATTGTCTGATTCGTTACCAATTATGACAACTTGACGGC

TACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTC

GGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAG

AGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATA

GGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATA

CGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGC

TGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAAC

ATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAG

GTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATC

CATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCG

CAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGA

ATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAA

CAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAA

GAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTC

ATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATA

CCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTC

TCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAA

TTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTG

AGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATA

AAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTTAAACCC

GCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGA

TCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCA

GAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTC

ACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGGTAAC

CCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGC

CATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAG

AAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTAT

GCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGAC

GCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTT

TGGGCTAGCAGGAGGAATTCACCATGGGTACTGCACTCGCAA

CGCTGGCTGGGAAGCTGGCTGAACGTGTCGGCATGGATTCTG

TCGACCCACAGGAACTGATCACCACTCTTCGCCAGACGGCAT

TTAAAGGTGATGCCAGCGATGCGCAGTTCATCGCATTACTGA

TCGTTGCCAACCAGTACGGCCTTAATCCGTGGACGAAAGAAA

TTTACGCCTTTCCTGATAAGCAGAATGGCATCGTTCCGGTGG

TGGGCGTTGATGGCTGGTCCCGCATCATCAATGAAAACCAGC

AGTTTGATGGCATGGACTTTGAGCAGGACAATGAATCCTGTA

CATGCCGGATTTACCGCAAGGACCGTAATCATCCGATCTGCG

TTACCGAATGGATGGATGAATGCCGCCGCGAACCATTCAAAA

CTCGCGAAGGCAGAGAAATCACGGGGCCGTGGCAGTCGCAT

CCCAAACGGATGTTACGTCATAAAGCCATGATTCAGTGTGCC

CGTCTGGCCTTCGGATTTGCTGGTATCTATGACAAGGATGAA

GCCGAGCGCATTGTCGAAAATACTGCATACACTGCAGAACGT

CAGCCGGAACGCGACATCACTCCGGTTAACGATGAAACCATG

CAGGAGATTAACACTCTGCTGATCGCCCTGGATAAAACATGG

GATGACGACTTATTGCCGCTCTGTTCCCAGATATTTCGCCGC

GACATTCGTGCATCGTCAGAACTGACACAGGCCGAAGCAGTA

AAAGCTCTTGGATTCCTGAAACAGAAAGCCGCAGAGCAGAAG

GTGGCAGCATGAAGTCGGTGTGGAGTGAAACGATGAAGAAAA

ATCGCGCTTTTTTGAAGTGGGCAGGGGGCAAGTATCCCCTGC

TTGATGATATTAAACGGCATTTGCCCAAGGGCGAATGTCTGG

TTGAGCCTTTTGTAGGTGCCGGGTCGGTGTTTCTCAACACCG

ACTTTTCTCGTTATATCCTTGCCGATATCAATAGCGACCTGAT

CAGTCTCTATAACATTGTGAAGATGCGTACTGATGAGTACGT

ACAGGCCGCACGCGAGCTGTTTGTTCCCGAAACAAATTGCGC

CGAGGTTTACTATCAGTTCCGCGAAGAGTTCAACAAAAGCCA

GGATCCGTTCCGTCGGGCGGTACTGTTTTTATATTTGAACCG

CTACGGTTACAACGGCCTGTGTCGTTACAATCTGCGCGGTGA

GTTTAACGTGCCGTTCGGCCGCTACAAAAAACCCTATTTCCC

GGAAGCAGAGTTGTATCACTTCGCTGAAAAAGCGCAGAATGC

CTTTTTCTATTGTGAGTCTTACGCCGATAGCATGGCGCGCGC

AGATGATGCATCCGTCGTCTATTGCGATCCGCCTTATGCACC

GCTGTCTGCGACCGCCAACTTTACGGCGTATCACACAAACAG

TTTTACGCTTGAACAACAAGCGCATCTGGCGGAGATCGCCGA

AGGTCTGGTTGAGCGCCATATTCCAGTGCTGATCTCCAATCA

CGATACGATGTTAACGCGTGAGTGGTATCAGCGCGCAAAATT

GCATGTCGTCAAAGTTCGACGCAGTATAAGCAGCAACGGCGG

CACACGTAAAAAGGTGGACGAACTGCTGGCTTTGTACAAACC

AGGAGTCGTTTCACCCGCGAAAAAATAATAAGCTTGGCTGTT

TTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAAT

C

pMAZ-Selfkilling (pMAZ-SK) sequence (20N is the location of the target sequence for the negative selection):

ATATACCATGGCCACAATTCAGCAAATTGTGAACATCATCACG

TTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTA

ACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTG

TTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACCCGTTCCG

TGTAGACAGTTCGCTCCAAGCTGTTTTAGAGCTATGCTGTTTT

GAATGGTCCCAAAACGCCATGTTTCAGAAACAACTCTGGCGC

ATCGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAAAA

AAAAACCCCGCCCCTGACAGGGCGGGGTTTTTTTTTTAATTAA

AGTCTGGTCCTCGAGTCTGGTTATAATCCCTATCAGTGATAGA

GATTGACATCCCTATCAGTGATAGAGATACTGAGCACNNNNN

NNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAA

AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG

TCGGTGCTTTTTTATGCAGCCTACCGGTATCCTAGGCTGCTGC

CACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAA

ACGGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGA

GAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGT

AAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCC

TGAACCGACGACAAGCTGACGACCGGGTCTCCGCAAGTGGCA

CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTT

CTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTA

GAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATA

TCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTA

ATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAA

GATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAA

TACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAG

TGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGG

CAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAG

CCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGT

TATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGGT

CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACC

GGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTG

AATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGG

GATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGAT

AAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCA

GTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTA

CCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTC

CCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTAT

CGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGA

ATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATG

GCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAG

GGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA

AAAATAAACAAATAGGCATGCTAGCGCAGAAACGTCCTAGAA

GATGCCAGGAGGATACTTAGCAGAGAGACAATAAGGCCGGA

GCGAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACGAACAT

CACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACA

GGACTATAAAGATACCAGGCGTTTCCCCCTGATGGCTCCCTC

TTGCGCTCTCCTGTTCCCGTCCTGCGGCGTCCGTGTTGTGGT

GGAGGCTTTACCCAAATCACCACGTCCCGTTCCGTGTAGACA

GTTCGCTCCAAGCTGGGCTGTGTGCAAGAACCCCCCGTTCAG

CCCGACTGCTGCGCCTTATCCGGTAACTATCATCTTGAGTCCA

ACCCGGAAAGACACGACAAAACGCCACTGGCAGCAGCCATTG

GTAACTGAGAATTAGTGGATTTAGATATCGAGAGTCTTGAAG

TGGTGGCCTAACAGAGGCTACACTGAAAGGACAGTATTTGGT

ATCTGCGCTCCACTAAAGCCAGTTACCAGGTTAAGCAGTTCC

CCAACTGACTTAACCTTCGATCAAACCGCCTCCCCAGGCGGT

TTTTTCGTTTACAGAGCAGGAGATTACGACGATCGTAAAAGG

ATCTCAAGAAGATCCTTTACGGATTCCCGACACCATCACTCTA

GATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTC

AGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCC

CCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAA

GGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACT

TACATTAATTGCGTTGCGCACACCGACTAGCGAAAAAACCCC

GCCGAAGCGGGGTTTTTTGCGAAAAAAAGCACCGACTCGGTG

CCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG

CTATGCTGTTTTGAATGGTTCCGCTAGCACTGTACCTAGGACT

GAGCTAGCCGTCAAAAGCAGCGT

Paper 2.

Ronda C., Maury J., Jakociunas T., Abdessamad Jacobsen S., Germann S.M., Harrison S.J., Borodina I., Keasling J.D., Jensen M.K., Nielsen A.T. **CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae.** *Microb. Cell Fact.* **14(1),** 97 (2015).

Ronda *et al. Microb Cell Fact (2015) 14:97*
DOI 10.1186/s12934-015-0288-3

MICROBIAL CELL FACTORIES

RESEARCH Open Access

CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*

Carlotta Ronda[†], Jérôme Maury[†], Tadas Jakočiūnas[†], Simo Abdessamad Baallal Jacobsen, Susanne Manuela Germann, Scott James Harrison, Irina Borodina, Jay D Keasling, Michael Krogh Jensen and Alex Toftgaard Nielsen[*]

Abstract

Background: One of the bottlenecks in production of biochemicals and pharmaceuticals in *Saccharomyces cerevisiae* is stable and homogeneous expression of pathway genes. Integration of genes into the genome of the production organism is often a preferred option when compared to expression from episomal vectors. Existing approaches for achieving stable simultaneous genome integrations of multiple DNA fragments often result in relatively low integration efficiencies and furthermore rely on the use of selection markers.

Results: Here, we have developed a novel method, CrEdit (CRISPR/Cas9 mediated genome Editing), which utilizes targeted double strand breaks caused by CRISPR/Cas9 to significantly increase the efficiency of homologous integration in order to edit and manipulate genomic DNA. Using CrEdit, the efficiency and locus specificity of targeted genome integrations reach close to 100% for single gene integration using short homology arms down to 60 base pairs both with and without selection. This enables direct and cost efficient inclusion of homology arms in PCR primers. As a proof of concept, a non-native β-carotene pathway was reconstructed in *S. cerevisiae* by simultaneous integration of three pathway genes into individual intergenic genomic sites. Using longer homology arms, we demonstrate highly efficient and locus-specific genome integration even without selection with up to 84% correct clones for simultaneous integration of three gene expression cassettes.

Conclusions: The CrEdit approach enables fast and cost effective genome integration for engineering of *S. cerevisiae*. Since the choice of the targeting sites is flexible, CrEdit is a powerful tool for diverse genome engineering applications.

Keywords: Metabolic engineering, CRISPR/Cas9, Genome editing, *Saccharomyces cerevisiae*, Carotenoid production, Genome integrations

Background

The production of bio-based chemicals, fuels, pharmaceuticals and food additives by microbial fermentation is a rapidly growing field. There is an increasing demand for efficient cell factories that enable the production of biofuels and biochemicals from renewable resources at low and competitive cost. The knowledge of genetics, physiology, biochemistry and large-scale fermentation of baker's yeast *Saccharomyces cerevisiae*, combined with the advent of genome engineering and recombinant DNA technology makes it a preferred host for many industrial bio-based applications, ranging from biofuels and bulk chemicals to nutraceuticals and pharmaceuticals [1–8]. Furthermore, *S. cerevisiae* has the advantage of being easy to manipulate genetically with a range of established cloning and vector systems [6, 9].

Production organisms with multi-enzyme pathways often require precise control of the expression level of the associated genes [2, 5, 10]. Besides regulating promoter strength, the copy number of genes is a critical control point. Both plasmid and genomic integration systems are

*Correspondence: atn@biosustain.dtu.dk
[†]Carlotta Ronda, Jérôme Maury and Tadas Jakočiūnas contributed equally to the work
The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark

widely used for heterologous expression of genes in *S. cerevisiae*. Plasmid-based systems typically offer limited control of copy number, and significant segregational instability of plasmids is often observed even during growth in selective medium [10]. It has for example been demonstrated that plasmid-based gene expression is highly heterogeneous, and that both 2μ and CEN/ARS vectors can be difficult to maintain at a stable level within the same cell population [11, 12]. Genomic integration is therefore the preferred alternative to ensure long-term stability and homogeneous expression of genes within a population.

Methods that enable fast, sequential or combinatorial integrations are valuable for metabolic engineering. Several powerful approaches, either plasmid- or PCR-based, have been demonstrated for genome integrations using selection markers. Such methods typically use active recombination systems, such as Cre/LoxP and FLP/FRT, to excise the marker without the need of counter selection [13, 14]. Recently, Jensen et al. developed an efficient set of vectors, the EasyClone vector set, that enables fast and simultaneous multiple integrations of genes into specific "safe sites of insertion" with the possibility of recycling the selective markers [12]. The insertion sites are located between essential elements, which limits the occurrence of chromosomal aberrations due to the lethal effect this would cause [15]. Based on homologous recombination using 500 bp long homology arms, this method results in successful integration into a single site [12]. However, the efficiency of integration decreases when native genes or promoters are present on the fragment to be integrated, or in the case of multiple simultaneous integrations (unpublished results). Jensen et al. reported 44% integration efficiency for simultaneous integration of three heterologous genes at three different loci using selection [12]. Increasing the efficiency of targeted integration without selection is therefore important for accelerating and potentially automating the strain engineering process.

The recent advent of CRISPR/Cas9 for genome engineering has enabled efficient genome editing in different organisms such as bacteria [16], mice [17], plants [18], fruit flies [19], fish [20] and mammalian cells [21–23]. CRISPR/Cas9 has also been applied for targeted single and multiple gene deletions in *S. cerevisiae* by homology-directed repair of double-strand breaks (DSBs) using short oligonucleotides as repair donors, in different strain backgrounds [24–29]. The prevalent DSB repair mechanism in *S. cerevisiae* is native homologous recombination (HR), and the introduction of a DSB has been shown to increase integration of heterologous linear DNA fragments with ends homologous to the DSB site [30, 31]. Harnessing HR for DSB repair, Ryan et al. recently reported the successful integration of a three-part DNA assembly into a single chromosomal locus [26], and Mans et al. performed a complete deletion of the *ACS2* locus in combination with a six-part DNA assembly that resulted in the deletion of the *ACS1* locus [26, 28]. This impressive approach, however, most likely requires additional intrinsic selection pressure, with the simultaneous deletion of these two loci being essential for viability. Furthermore both Horwitz et al. and Jakociunas et al. have recently shown multiplex assembly and integration of multiple parts in three loci, albeit with relatively low efficiencies [27, 29, 32]. Jakociunas et al. have demonstrated the powerful application of the CRISPR/Cas9 system as a tool for metabolic engineering utilizing user-friendly and easy-to-use USER-technology-based gRNA constructs [27]. In order to further expand this existing platform for knock out constructions, we wished to investigate whether CRISPR/Cas9 together with the DNA brick based EasyClone approach could be employed for targeted one-step selection-free integration of multiple genes into the *S. cerevisiae* genome.

Here, we have developed a system, CrEdit (CRISPR/Cas9 mediated genome Editing), which combines the high specificity of CRISPR/Cas9 with the convenient genome engineering tool EasyClone for achieving highly efficient and accurate simultaneous genomic integration of multiple pathway gene expression cassettes in different loci in the genome of *S. cerevisiae*. The gRNA-guided Cas9 endonuclease was used to target gene integration at selected insertion sites, which resulted in up to 100% correct selection-free target integration at the desired locus for the donor DNA. CrEdit also enabled simultaneous and highly efficient integration of three pathway genes involved in the production of β-carotene at three different integration sites located on three different chromosomes.

Results and discussion

Construction of the CrEdit system

In order to increase the efficiency of targeted integration into the *S. cerevisiae* genome, we decided to combine the well-characterized genomic integration sites used in the EasyClone system with the RNA-guided endonuclease activity of Cas9. Initially, we tested two different designs for the system. In the first design, Cas9 was expressed from a constitutive promoter, P_{TEF1}, on an ARS/CEN based vector, while the gRNA that targets Cas9 to the chosen EasyClone integration site was expressed from an episomal 2μ-based vector (Figure 1) [24]. In the second design, Cas9 was under the control of the inducible P_{CUP1} promoter and integrated in the genome, and the gRNA supplied on a linearized integrative vector. The first design was chosen for its versatile and recyclable aspects, while the second design was chosen for the possibility of

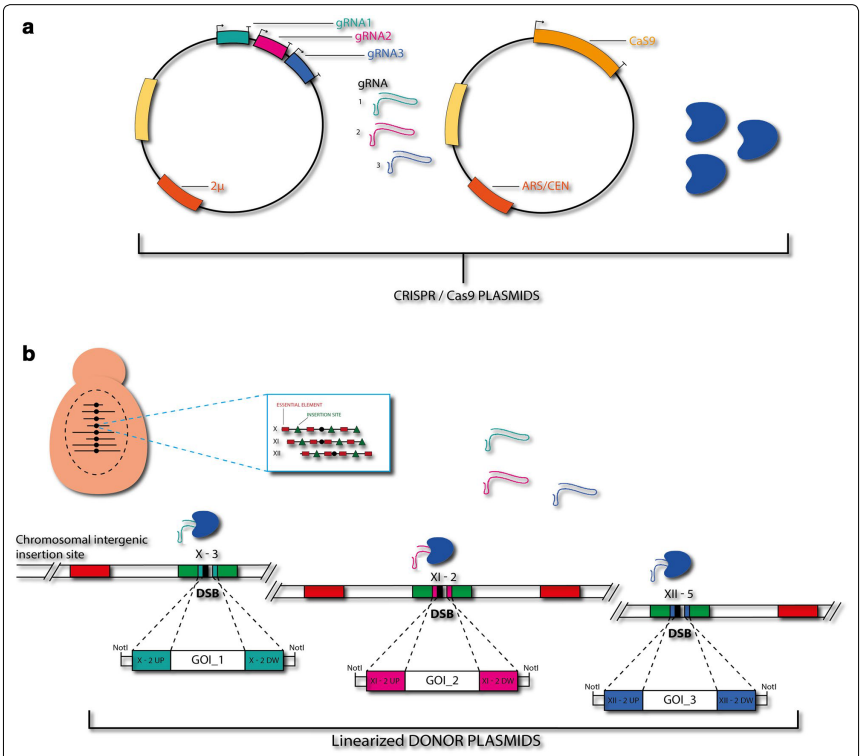


Figure 1 Schematic overview of the CrEdit system. **a** Replicative vectors expressing gRNAs and *cas9*. **b** Targeted DSBs mediated by Cas9 endonuclease activity facilitate the integration of linearized donor plasmids by homologous recombination. Donor plasmids harbor the desired integration sequences flanked by homology arms. Selected intergenic safe harbor sites can be used for simultaneous integration of multiple genomic expression cassettes for pathway engineering.

controlling the expression of Cas9 and gRNAs at lower levels. Both types of gRNA carrier plasmids have been designed to enable a fast exchange of the gRNA expression cassettes via USER cloning. Thereby, it is possible to conveniently target a new locus by quick and easy single-step cloning of the gRNA plasmids [27]. Also, the USER-overhang system enables multiplexing of up to five gRNAs on one single plasmid [27].

In this study, we show the use of both genomic and plasmid versions in combination with the donor DNA being provided via EasyClone integration plasmids. The donor DNA can contain up to two promoter-gene-terminator sequences, a selection marker flanked with loxP sites, and homology arms for homologous recombination at the defined insertion sites of the EasyClone system [12]. Importantly, for targeting integration site X-3, the sequence of the donor integration plasmid was modified by eliminating the PAM site (protospacer-adjacent motif, i.e. three nucleotides necessary for Cas9 recognition), since the PAM is located on a donor homology arm. This

design prevents Cas9 from cutting the target sequence once successful integration has occurred. In the other sites used, the PAM site is located within a section of the genome that is deleted by the successful integration event of the two interspaced homology arms. Since the PAM sequence is removed in case of completed integration, this might have an additional positive effect on obtaining correct transformants, since Cas9 keeps cutting in cells where integration was not successful. Thereby the DSB fails to be repaired, which is lethal for the cells [24].

Targeted single genomic integration of *tHMG1*

As a proof of concept for the applicability of CrEdit for metabolic engineering, we used the well-established carotenoid biosynthetic pathway as a model. Carotenoids are part of the diverse group of natural compounds called isoprenoids, and are synthesized from precursors derived from the native mevalonic acid (MVA) pathway (Figure 2). The *tHMG1* gene encodes a truncated HMG-CoA reductase, which has been shown to increase carbon flux through the pathway, leading to increased isoprenoid and carotenoid production [33, 34]. Therefore, we initially focused on introducing one copy of the *tHMG1* overexpression cassette into the *S. cerevisiae* genome.

In order to test the efficiency of the two different CrEdit designs, we decided to test single integration of donor DNA with differently sized homology arms. As donor we used an EasyClone integrative plasmid containing *tHMG1* with homology arms specific for intergenic site X-2 (Figure 3a) [15]. The integration efficiencies of all experiments are shown in Additional file 1: Table S1. We first tested the integration efficiency of using integrative gRNA in combination with a *S. cerevisiae* strain harboring genomic *Cas9* under the control of the P_{CUP1} promoter. Cas9 expression was induced by addition of Cu^{2+} 2 h before transformation. We then co-transformed this Cas9-expressing strain with the specific donor DNA carrying *tHMG1* with homology arms of 500, 110 or 60 bp length for site X-2, and the integrative gRNA targeting site X-2. An empty vector backbone without gRNA was used as a control. The resulting transformants were plated onto medium selecting for Cas9, the gRNA and the donor selection marker. We then analyzed the genotype of at least 16 colonies per condition to check for correct insertion at site X-2. When relying solely on intrinsic homologous recombination, the measured efficiency of correct integration at site X-2 was 70% with homology arms of approximately 500 bp (Figure 3b, left panel, −gRNA). As expected, the efficiency of correct integration was found to decrease significantly when using shorter arms with lengths of either 110 or 60 bp (Figure 3b, left panel, −gRNA). However, when the gRNA targeting X-2 was expressed, close to 100%

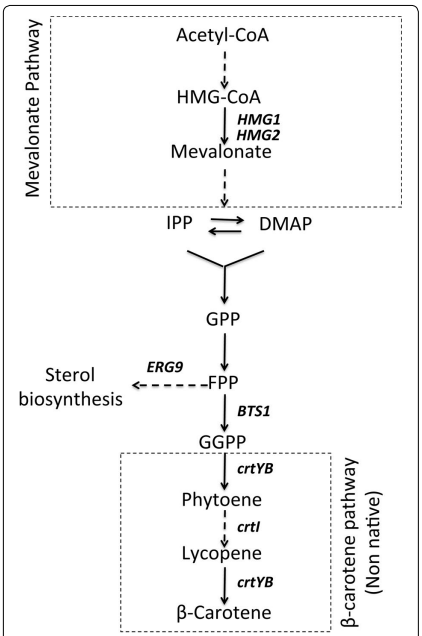


Figure 2 Overview of the biosynthetic pathway for β-carotene production. The carotenoid biosynthetic pathway can be reconstructed in *S. cerevisiae* by overexpression of the native GGPP synthase encoded by *BTS1*, and co-overexpression of the non-native bifunctional phytoene synthase/lycopene cyclase encoded by *crtYB*, and phytoene desaturase encoded by *crtI* of *X. dendrorhous*. *HMG1* encodes the major HMG-CoA reductase activity in *S. cerevisiae*. *ERG9* encodes a farnesyl-diphosphate farnesyl transferase (squalene synthase) that acts in the sterol biosynthesis pathway. *IPP* isopentenyl diphosphate, *DMAP* dimethylallyl diphosphate, *GPP* geranyl diphosphate, *FPP* farnesyl diphosphate, *GGPP* geranylgeranyl diphosphate.

successful integration was obtained at site X-2, regardless of the length of the homology arms (Figure 3b, left panel, +gRNA). Interestingly, when using the plasmid-based gRNA/Cas9 system and in the absence of gRNA, 100% correct integrants could only be obtained using 500 bp homology arms. Furthermore, and only in that condition, a low number of transformants was obtained on plates, which points towards a negative effect of *cas9* expression on cells when expressed from the constitutive strong *TEF1* promoter and in the absence of gRNA. Ryan et al.

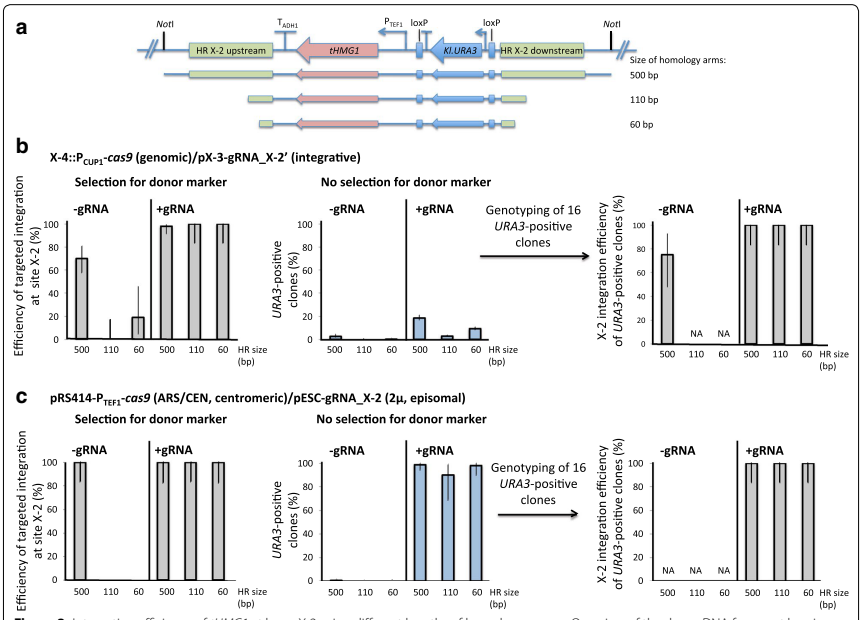


Figure 3 Integration efficiency of *tHMG1* at locus X-2 using different lengths of homology arms. **a** Overview of the donor DNA fragment bearing *tHMG1* with differently sized homology arms. **b** Integration efficiency of the CrEdit system with genomic inducible Cas9 and integrative gRNA. *S. cerevisiae* strain ST1011 harboring P_{CUP1}-*cas9* was induced with Cu^{2+} 2 h prior to transformation start, and then co-transformed with (*left*, −gRNA) linearized empty vector pCfB257 and linearized donor DNA encoding *tHMG1* (for details of donor DNA see Additional file 1), or (*right*, +gRNA) the linearized integrative gRNA vector pCfB2831 targeting X-2 and linearized donor DNA encoding *tHMG1*. *Left panel* Efficiency of targeted integration at site X-2 when selecting for donor DNA after transformation. *Middle panel* Efficiency of marker gene integration when not selecting for donor DNA after transformation. *Right panel* Frequency of correct integration at site X-2 determined by genotyping of URA^+ colonies. **c** Integration efficiency of the CrEdit system with plasmid-based Cas9 and gRNA. *S. cerevisiae* strain TC-3 harboring P_{TEF1}-*cas9* on the centromeric plasmid pCfB1767 was co-transformed with (*left*, −gRNA) empty vector pCfB2999 and linearized donor DNA encoding *tHMG1*, or (*right*, +gRNA) the episomal gRNA vector pCfB3020 targeting X-2 and linearized donor DNA encoding *tHMG1*. *Left panel* Efficiency of targeted integration at site X-2 when selecting for donor DNA after transformation. *Middle panel* Efficiency of marker gene integration when not selecting for donor DNA after transformation. *Right panel* Frequency of correct integration at site X-2 determined by genotyping of URA^+ colonies. Only +gRNA colonies were analyzed since no URA^+ clones were obtained in the absence of gRNA. The experiment was repeated twice and error bars represent 95% confidence intervals. *NA* not analyzed.

reported a decreased fitness of yeast strains expressing *cas9* from the strong *TDH3* promoter [26], while Mans et al. reported that the constitutive expression of *cas9* from the genome and the *TEF1* promoter does not affect the maximal specific growth rate on glucose based synthetic media [28]. In light of these results, a more detailed study of the impact of *cas9* expression levels on yeast cell physiology and especially the HR machinery is of interest. Still, 100% correct integrants were obtained in the presence of gRNA for all sizes of homology arms (Figure 3c, left panel), suggesting that the plasmid-based gRNA/Cas9 system also is very efficient.

In conclusion, we show that the DSB created by the guide RNA-targeted Cas9 endonuclease is instrumental for correct integration at a significantly higher efficiency than what can be achieved solely by endogenous homologous recombination. The lower efficiency observed in absence of CRISPR/Cas9 is possibly due to the fact that native genes tend to recombine at the native locus due to the large homology region. Also, expression cassettes

might integrate elsewhere in the genome possibly via break-induced replication (BIR), thus creating strains where it becomes difficult to localize the gene of interest. The targeted DSB created by Cas9 likely boosts HR at the desired integration site.

Targeted genomic integration without selective pressure

Because of the high efficiency observed for integration of *tHMG1*, we investigated if integration of this gene expression cassette could be performed even without applying selection pressure for the donor DNA marker *Kl.URA3*. We repeated the integration experiment described above, however this time plating the transformants on medium only selecting for gRNA and Cas9. When the plasmid-based gRNA/Cas9 CrEdit system was used, 99, 90, and 98% efficiency of integration of the marker gene was observed for 500, 110 and 60 bp homology arms, respectively (Figure 3c, middle panel, +gRNA). The PCR analysis at locus X-2 for the resulting *Kl.URA3*-positive clones showed 100% correct integration into site X-2 for all tested sizes of homology arms (Figure 3c, right panel, +gRNA). However, when using the genomic CrEdit system with induced P_{CUP1}-*cas9*, only 19, 3 and 9% integration efficiency were achieved for 500, 110 and 60 bp homology arms, respectively (Figure 3b, middle panel, +gRNA). Despite the lower integration efficiency, PCR analysis of the resulting *Kl.URA3*-positive clones showed 100% correct integration into site X-2 for all tested sizes of homology arms (Figure 3b, right panel, +gRNA). When the empty vector (−gRNA) was included in the transformation, the efficiency of marker integration was close to zero in all cases, independent on the length of the homology arms (Figure 3b, c, middle panels, −gRNA). In the case of genomic *cas9* and long 500 bp homology arms, the genotyping of 16 *Kl.URA3*-positive clones showed approximately 75% correct integration at site X-2 (Figure 3b, right panel, −gRNA). Differences in promoters between the systems, and the time-limited induction of *cas9* by the *CUP1* promoter in our experimental set-up (2 h prior to transformation) may lead to lower levels of Cas9 at transformation start compared to the plasmid-based system where *cas9* is under the control of the constitutive *TEF1* promoter on a centromeric plasmid. In conclusion, the highest efficiency of both selection- and non-selection based genomic integration was achieved when both gRNA and *cas9* were expressed from plasmids, and we therefore chose this to be the final configuration of the CrEdit system (Figure 1).

Targeted simultaneous multi-loci integration of three carotenogenic pathway genes

In order to speed up the strain construction process, it is often desirable to simultaneously insert multiple genes into the genome. After having achieved highly efficient insertion of *tHMG1* into intergenic site X-2 using the CrEdit method, we tested simultaneous integration of multiple genes into the genome of *S. cerevisiae*. As a proof of concept, we attempted to introduce the non-native production of carotenoids in *S. cerevisiae* via expression of the two heterologous genes *crtYB* and *crtI* of *X. dendrorhous* combined with overexpression of *S. cerevisiae* geranylgeranyl diphosphate (GGPP) synthase encoded by *BTS1* [35]. The gene *crtYB* encodes a bifunctional enzyme with phytoene synthase and lycopene cyclase activity, while *crtI* encodes a phytoene desaturase [36].

Using the plasmid-based CrEdit system, cells expressing Cas9 were simultaneously transformed with three different large EasyClone donor DNAs for integration of P_{TDH3}-*crtI* (6.6 kb), P_{TEF1}-*crtYB* (5.8 kb), and P_{PGK1}-*BTS1* (5.1 kb) into three intergenic sites X-3, XI-2, and XII-5 situated on different chromosomes, using 500 bp homology arms. The cells were co-transformed with one episomal vector expressing the three gRNAs targeting these three sites, or with the empty vector for the −gRNA control. Transformants were plated on media selecting only for Cas9 and gRNA expressing plasmids. We observed that 84% of the derived colonies presented orange pigment formation when the gRNAs were present, indicating complete β-carotene pathway integration. In contrast, we only observed white colonies when the gRNAs were absent, indicating that no correct triple integration had been achieved (Figure 4a). All colonies were then replicated on single drop-out plates in order to screen for the integration of the three independent selection marker genes. As expected, all orange colonies were positive for all the three marker genes (Figure 4b, left panel). We subsequently tested the genotype of 32 orange colonies at the three expected integration sites, and observed 100% correct integration, thereby confirming complete pathway assembly (Figure 4b, right panel). In addition, we measured β-carotene levels by HPLC in three confirmed clones, and demonstrated that 12.7 ± 2.5 mg L^{-1} β-carotene was produced (Figure 4c). This proves the ability of the CrEdit system to simultaneously integrate three large DNA fragments with surprisingly high efficiency (84%) at the correct loci even without selection pressure. As for comparison, simultaneous integration of three genes has previously been demonstrated with 44% efficiency when relying on native HR alone and when applying selective pressure [12]. It was furthermore attempted to repeat the multi-loci pathway integration using short homology arms (60 bp) to investigate if we could simply use PCR products directly as donors for the multiplex integration. However, no viable colonies grew on the plates even after 1 week of incubation (data not

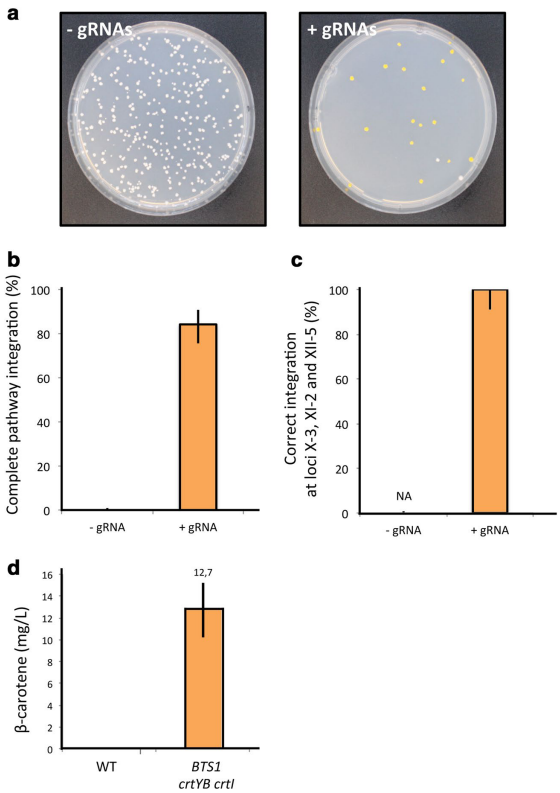


Figure 4 Multiple simultaneous integration of three β-carotene pathway genes. **a** Picture of colonies after simultaneous integration of *BTS1*, *crtYB* and *crtI* on plates without selection. *S. cerevisiae* (TC-3) harboring a centromeric plasmid constitutively expressing *cas9* was co-transformed with: *left* empty vector control and linearized donor DNAs encoding *BTS1*, *crtYB* and *crtI*. *Right* gRNA vector expressing three gRNAs targeting intergenic loci XII-5, XI-2 and X-3, and linearized donor DNAs encoding *BTS1*, *crtYB* and *crtI*. Colonies with successful pathway integration accumulate β-carotene, resulting in an orange pigmentation. **b** Percentage of complete pathway integration with and without the expression of targeting gRNAs. **c** Frequency of correct simultaneous integration of the three genes *BTS1*, *crtYB*, and *crtI* at the specific intergenic loci XII-5, XI-2 and X-3, respectively, determined by genotyping. **d** HPLC analysis of β-carotene production of three independent orange colonies and a non-producing strain as control (CEN.PK113-7D). The experiment was repeated twice and *error bars* in all *panels* represent 95% confidence intervals.

shown), indicating that longer homology arms are beneficial for multiplex genome integrations. We assume this may due to the fact that multi-loci pathway integration is quite demanding with regards to coordinated repair activity, and long homology arms are easier to utilize for the native yeast HR machinery, thereby enabling correct simultaneous integration at multiple loci.

The results obtained for simultaneous integration of three genes (*BTS1*, *crtYB* and *crtI*) show the ability of the CrEdit system to insert very large fragments (up to

17.5 kb in this study) without the need for a selection marker, which is very attractive for industrial metabolic engineering applications. Industrial strains are often prototrophic, and/or diploid or even polyploid, thus making the use of auxotrophic markers challenging. Furthermore, even for haploid auxotrophic strains, the limited number of available selection markers typically necessitates recycling of the markers. Several systems can be used for looping out genetic elements, including the Cre-LoxP and FRT/FLP systems [12–14]. Such methods are not only time consuming but can also leave scars, which can cause genome instability and rearrangements in recombinant strains [37]. Importantly, CrEdit enables selection-free and scarless integration of desired DNA sequences, thereby limiting the risk for strain instability while significantly speeding up strain engineering. Moreover, CrEdit is a versatile genome engineering tool, since the design of novel gRNAs for alternative integration sites can be easily achieved using for example the recently developed in silico gRNA selection tool, CRISPy, which minimizes the potential risk of off-target effects of Cas9 activity [23, 27].

Recently, high efficiencies using CRISPR/Cas9 system for genome integration have been reported, but most systems either still rely on selective pressure or, if selection was not applied, only short DNA sequences were inserted [25, 26]. In the recent work by Horwitz et al., an 11 gene pathway was integrated via 6-part integration, however only very low efficiency was observed [32]. Stovicek et al. also demonstrated successful multi-part assembly at three different loci, yet with relatively low efficiencies [29]. In contrast, CrEdit is a versatile system for achieving high efficiency of single and multiple simultaneous integrations without the need for selection (when long homology arms are used). The CrEdit system was designed in a way that the PAM sequence is eliminated upon successful integration. The continued cutting of the wild-type DNA is thereby possibly contributing to the very high efficiency of integration [24].

Further engineering of carotenoid production

A significant amount of work is available on engineering organisms for production of carotenoids [38–41], and in recent years a biosustainable and economically attractive production of β-carotene has been achieved [34, 42]. In an attempt to further boost β-carotene production, we integrated the β-carotene pathway in *S. cerevisiae* strains bearing genetic modifications previously reported as being beneficial for the flux to the mevalonate pathway. We therefore performed the multi-loci integration experiment in a CEN.PK strain carrying a down-regulated version of the squalene synthase *ERG9* gene (*erg9*::Δ-220–176). In this strain, a deletion of an upstream section of the promoter causes lower *ERG9* transcript and protein levels, thereby reducing the flux towards the competing endogenous sterol biosynthetic pathway [27, 43]. We also transformed a CEN.PK strain that carried both the *erg9*::Δ-220–176 modification and an overexpression of *tHMG1*. Orange colonies producing carotenoids were obtained with high efficiency in both genetic backgrounds (Additional file 1: Figure S1). It was also observed that these latter strains were clearly less orange compared to the unmodified CEN.PK strain only expressing the β-carotene pathway (compare Figure 4 and Additional file 1: Figure S1). β-carotene concentrations were measured and it was shown that the additional genetic modifications did not lead to an increase in the β-carotene levels. A significant decrease in β-carotene concentration was even observed for the strain bearing both *erg9*::Δ-220–176 and the overexpression of *tHMG1* (Additional file 1: Figure S1). Indeed, Verwaal et al. have shown that the desaturation of phytoene, catalyzed by CrtI, is a rate-limiting step in carotenoid production, and that an increase of the total carotenoid accumulation is largely caused by a significant increase of this precursor [35]. As phytoene is color-less, it is expected that its accumulation in the strains improved for precursor availability results in the less intense coloration of the yeast colonies. In order to avoid this precursor accumulation, it may be possible to further boost the expression of *crtI* by integrating this pathway gene in more copies [35].

Conclusion

In summary, we were able to demonstrate the ability of the CrEdit system to simultaneously integrate up to three large DNA fragments with high efficiency even without selective pressure into different genetic backgrounds, supporting the strength and robustness of the method.

CrEdit combines the stability and versatility of the EasyClone vector system with the precision and efficiency of CRISPR/Cas9, thereby significantly increasing the efficiency of genome integrations in *S. cerevisiae*. We demonstrate how this system can be used for simultaneous integration of multiple genes with high efficiency, even without selection for donor DNA. CrEdit is also very efficient in integrating large fragments at single loci using short homology arms of 60 bp that can be included in PCR primers. This facilitates quick and easy exchange from one integration site to another. A further advantage of the primer-based preparation of donor DNA is that the PAM recognition site can easily be removed from the short homology arms. Provided that a suitable PAM sequence is present at the genomic site of interest, the system can easily be developed for other genome engineering applications, such as combining integrations with gene deletions, defined site-specific mutagenesis,

gene replacements, promoter exchange, protein domain swapping, in a scarless and selection-free manner. We therefore believe that CrEdit will be a valuable genome engineering tool to facilitate fast and cost-effective production strain engineering.

Methods

Strains and media

Saccharomyces cerevisiae CEN.PK strains were obtained from Peter Kötter (Johann Wolfgang Goethe-University Frankfurt, Germany). All yeast strains used in this study were derivatives of CEN.PK (Additional file 1: Table S2). All standard cloning was carried out using *E. coli* strain DH5alpha. Media and standard genetic techniques used for manipulating yeast strains were performed as previously described [44]. Synthetic complete medium as well as drop-out media and agar plates were prepared using premixed drop-out powders (Sigma-Aldrich). All chemicals were obtained from Sigma-Aldrich. *Escherichia coli* transformants were grown in standard Luria–Bertani (LB) medium containing 100 μg mL^{-1} ampicillin.

Construction of plasmids for single targeted integration

All plasmids are described in Additional file 1: Table S3, and all gRNA sequences are listed in the Additional file 1 as well. Construction of expression plasmids used as donor DNA for integration is explained in detail in the Additional file 1. For design of all gRNA target sequences, the overall design was based on DiCarlo et al. [24] (Additional file 1: Table S4), and for designing the target sequence the program CRISPy was used [23, 27]. gRNA plasmid pCfB2831 used for integrating the gRNA X-2′ (targeting site X-2) into chromosomal site X-3 [15] was constructed by amplifying a gRNA expression cassette (ordered from Integrated DNA Technologies as gBlock), gRNA_X-2′ (Additional file 1: Table S5), with primers PR-10735/PR-10736 (Additional file 1: Table S6), and subsequent USER cloning into *Asi*SI/*Nb.Bsm*I-digested pCfB257 according to Jensen et al. [12]. To construct the episomal gRNA plasmid pTAJAK-76 (targeting site X-2), a backbone-cloning vector was created for USER cloning of the gRNA expression cassettes by amplifying and re-ligating pESC-LEU with TJOS-97F and TJOS-97R. Secondly, the resulting vector was amplified using the primers TJOS-108 and TJOS-102R in order to remove the *KlLEU2* marker. The NatMXsyn marker was then amplified from pCfB2180 (GeneArt) with the primers TJOS-106F and TJOS-106R, and USER-cloned into the vector lacking the *KlLEU2* marker, resulting in plasmid pTAJAK-71. Finally, to target site X-2 [15] with Cas9, a gRNA expression cassette was ordered from Integrated DNA Technologies as gBlock, gRNA_X-2 (Additional file 1: Table S5), and amplified with following primers: TJOS-62, TJOS-65. Amplified gRNA was USER cloned into pTAJAK-71, which was previously digested with *Asi*SI/*Nb.Bsm*I, resulting in the plasmid pTAJAK-76.

Construction of plasmids carrying multiple gRNAs

First, a backbone-cloning vector was created for USER cloning of the gRNA expression cassettes by amplifying and re-ligating pESC-LEU with TJOS-97F and TJOS-97R. Secondly, the resulting vector was amplified using the primers TJOS-108 and TJOS-102R in order to remove the *KlLEU2* marker. The KanMXsyn marker was then amplified from pCfB2179 (GeneArt) with the primers TJOS-106F and TJOS-106R, and cloned into the vector lacking the *KlLEU2* marker, resulting in plasmid pTAJAK-72. Finally, to target the sites X-3, XI-2 and XII-5 [15] with Cas9, gRNA expression cassettes [24] were ordered from Integrated DNA Technologies as gBlocks (gRNA sequences are listed in Additional file 1: Table S4) and amplified with following primers: TJOS-62/TJOS-66 (gRNA_X-3); TJOS-63/TJOS-67 (gRNA_XI-2) and TJOS-64/TJOS-65 (gRNA_XII-5). Amplified gRNAs were USER cloned into pTAJAK-72, which was previously digested with *Asi*SI/*Nb.Bsm*I, resulting in the plasmid pTAJAK-92 according to Ref. [27].

Transformation protocol for single integration

Plasmids were transformed into *S. cerevisiae* cells using the lithium acetate transformation protocol [45]. Initially, P_{CUP1}-*cas9* was integrated into EasyClone site X-4 by transforming CEN.PK102-5B with the integrative vector pCFB1129 resulting in strain ST1011 (Additional file 1: Table S3). When transforming strain ST1011, Cas9 transcription was induced by adding 200 μM $CuSO_4$ 2 h before harvesting the cells for transformation. Prior to transformation, donor DNA was prepared as follows. For 500 bp homology arms, the integrative vector pCfB772 was digested by *Not*I and column-purified (Nucelospin Gel and PCR cleanup kit, Macherey Nagel). For shorter homology arms, pCfB772 was amplified by PCR using primer sets PR-9706/PR-9707 (110 bp) or PR-9704/PR-9705 (60 bp), *Dpn*I-treated and resolved on 1% agarose gel containing SYBR®-SAFE (Invitrogen) and purified using NucleoSpin® Gel and PCR Clean-up kit (Macherey Nagel). For single integration, 1 μg donor DNA and 1 μg *Not*I-linearized integrative gRNA plasmid or 500 ng undigested episomal gRNA plasmid was co-transformed into competent yeast cells. Cells were plated on media that selected for the presence of the gRNA (*KlLEU2*) and Cas9 (*SpHIS5*), and optionally donor marker (*KlURA3*) where stated. When colonies appeared, the transformation plates were replicated on selective plates (SC-LEU, SC-URA, SC-HIS) to screen for colonies with integrated selection markers. Correct integration at the specific

genomic locus was verified by colony PCR with following primers: PR-2221/PR-901 (X-2: P_{TEF1}-*tHMG1*).

Transformation protocol for multiple integration of carotenoid pathway

To simultaneously integrate three genes required for carotene production, 3 μg of each carrier plasmid (pTAJAK-94, pTAJAK-95, pTAJAK-12) were linearized by *NotI* digestion. *S. cerevisiae* strain TC-3 [27] was co-transformed with these linearized donor plasmids plus 1 μg of triple gRNA plasmid pTAJAK-92. Cells were plated on media that selected for the presence of the gRNA (kanMX) and Cas9 (*TRP1*) plasmids. When colonies appeared, the transformation plates were replicated on selective plates (SC-LEU, SC-URA, SC-HIS) to screen for colonies with integrated selection markers. To screen for correct integrations to the expected loci of carotene genes, colony PCR was performed with following primers: PR-2221/PR-903 (X-3: P_{TDH3}-*crtI*); PR-2221/PR-909 (XI-2: P_{TEF1}-*crtYB*); PR-2221/PR-899 (XII-5: P_{PGK1}-*BTS1*). The experiment was carried out in triplicate, and statistical analysis (one-tailed Student's *t* test) was performed on the complete data set. Multiple integration of carotenoid pathway was further performed in strains TC-23 and ST3450, according to the protocol mentioned above. Strain TC-23 harbors a *erg*9::Δ-220–176 genetic modification [30]. Strain ST3450 was obtained by transforming *S. cerevisiae* strain TC-23 with a *Not*I linearized pCfB2996 and transformants were selected on medium containing nourseothricin. Strain ST3450 therefore harbors *erg*9::Δ-220–176 and a copy of P_{TEF1}-t*HMG1* integrated at chromosome locus X-2.

β-Carotene quantification

Three independent orange colonies from *S. cerevisiae* TC-3 containing the three expression cassettes for *BTS1*, *crtYB and crtI* were used to inoculate test tubes containing 4 mL of drop out medium per well. As a reference, a colony of *S. cerevisiae* CEN. PK 113-7D was inoculated in the same conditions, and all cells were cultivated at 30°C with 300 r.p.m shaking. After approximately 48 h of cultivation, 3.5 mL of cultivation broth was centrifuged for 5 min at 4,000 rpm. Then supernatants were discarded and cell pellets resuspended in 0.2 mL of milliQ water. Cell suspensions were transferred to screw-cap tubes, suitable for subsequent cell breakage in a Precellys homogenizer. Glass beads and 1 mL of hexane were added to the cell suspension and cells were mechanically lysed for four cycles, each of 20 s at 6,500 rpm. Tubes were placed on ice for 1 min in between each lysis cycle. Subsequently, tubes were centrifuged for 5 min at 10,000 rpm to separate cell debris, aqueous and solvent fractions. The hexane fraction was collected in glass vials. Hexane was then evaporated in a rotary evaporator, under vacuum, and dry extracts were re-dissolved in 1 mL of ethanol 99%. Extracts were then analysed by LC–MS. LC–MS data was collected on Orbitrap Fusion equipped with a Dionex brand Ultimate 3000 UHPLC pumping system (ThermoFisher Scientific, Waltham, MA, USA). Samples were held in the autosampler at a temperature of 10.0°C during the analysis. 2 μL injections of the sample were made onto a Supelco Discovery HS F5-3 HPLC column, with a 3 μm particle size, 2.1 mm i.d. and 150 mm long. The column was held at a temperature of 30.0°C. The solvent system used was Solvent A "Water with 0.1% formic acid" and Solvent B "Acetonitrile with 0.1% formic acid". The flow rate was 1.000 mL min^{-1} with an initial solvent composition of %A = 75, %B = 25.0 held until 3.0 min, the solvent composition was then changed following a linear gradient until it reached %A = 0.0 and %B = 100.0 at 15.0 min. This was continued until 20 min, when the solvent was returned to the initial conditions and the column was re-equilibrated until 25 min. The column eluent flowed directly into the Heated ESI probe of the MS which was held at 325°C and a voltage of 3,500 V. Profile data was collected in positive ion mode with resolution setting of 30K and scan range (m/z) = 50–600. The other MS settings were as follows, sheath gas flow rate of 60 units, Aux gas flow rate of 20 units, sweep gas flow rate of 5 units, ion transfer tube temp was 380°C, maximum injection time of 100 ms, S-lens RF level = 60 V, using 1 Microscans and AGC target = 200,000 counts.

Additional file

Additional file 1: Figure S1. Showing efficiency of single step integration of the beta-carotenoid pathway in different strain backgrounds. Additionally, the file includes supplementary methods. **Table S1**. Efficiency of targeted integration using CrEdit. **Table S2**. List of strains used. **Table S3**. List of plasmids used. **Table S4**. gRNA sequences. **Table S5**. DNA and BioBricks and gBlocks. **Table S6**. Primers used in this study.

Abbreviations

CRISPR: clustered regularly interspaced short palindromic repeats; PAM: protospacer adjacent motif; FRT: flippase recognition target; FLP: flippase; DSB: double strand break; HR: homologous recombination; gRNA: guide RNA; KI.URA3: *Kluyveromyces lactis URA3* gene; Cre: cyclization recombinase.

Authors' contributions

ATN, CR, JM, IB, JDK and MKJ conceived the study; CR, JM, TJ, SAJ, and SMG performed the experiments and analyzed the data; CR, TJ, SAJ and SMG generated plasmids; JM carried out cultivation and extraction experiments; SH performed metabolite analysis; CR drafted the manuscript and all the authors contributed to preparing the final version of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was financed by The Novo Nordisk Foundation. We thank Prof. Gerhard Sandmann for plasmids containing carotene genes, and Mette Kristensen for analytics assistance.

Compliance with ethical guidelines

Competing interests
The authors declare that they have no competing interests.

Received: 8 February 2015 Accepted: 22 June 2015
Published online: 07 July 2015

References

1. Nielsen J, Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res 8:122–131
2. Borodina I, Nielsen J (2014) Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. Biotechnol J 9:609–620
3. Hong K-K, Nielsen J (2012) Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. Cell Mol Life Sci 69:2671–2690
4. Li M, Borodina I (2015) Application of synthetic biology for production of chemicals in yeast *Saccharomyces cerevisiae*. FEMS Yeast Res 15(1):1–12. doi:10.1111/1567-1364.12213
5. Tippmann S, Chen Y, Siewers V, Nielsen J (2013) From flavors and pharmaceuticals to advanced biofuels: production of isoprenoids in *Saccharomyces cerevisiae*. Biotechnol J 8:1435–1444
6. Krivoruchko A, Siewers V, Nielsen J (2011) Opportunities for yeast metabolic engineering: lessons from synthetic biology. Biotechnol J 6:262–276
7. Jensen MK, Keasling JD (2015) Recent applications of synthetic biology tools for yeast metabolic engineering. FEMS Yeast Res 15(1):1–10. doi:10.1111/1567-1364.12185
8. Mora-Pale M, Sanchez-Rodriguez SP, Linhardt RJ, Dordick JS, Koffas MAG (2014) Biochemical strategies for enhancing the in vivo production of natural products with pharmaceutical potential. Curr Opin Biotechnol 25:86–94
9. David F, Siewers V (2015) Advances in yeast genome engineering. FEMS Yeast Res 15(1):1–14. doi:10.1111/1567-1364.12200
10. Da Silva NA, Srikrishnan S (2012) Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. FEMS Yeast Res 12:197–214
11. Futcher B, Carbon J (1986) Toxic effects of excess cloned centromeres. Mol Cell Biol 6:2213–2222
12. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH et al (2014) EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Res 14:238–248
13. Johansson B, Hahn-Hägerdal B (2002) Overproduction of pentose phosphate pathway enzymes using a new CRE-loxP expression vector for repeated genomic integration in *Saccharomyces cerevisiae*. Yeast 19:225–231
14. Radhakrishnan P, Srivastava V (2005) Utility of the FLP-FRT recombination system for genetic manipulation of rice. Plant Cell Rep 23:721–726
15. Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH et al (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng 14:104–111
16. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233–239
17. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F et al (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153:910–918
18. Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41:e188
19. Bassett AR, Tibbit C, Ponting CP, Liu J-L (2013) Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4:220–228
20. Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X et al (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 23:465–472
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826
22. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823
23. Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT et al (2014) Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol Bioeng 111:1604–1616
24. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343
25. Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N et al (2014) Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. ACS Synth Biol 4(5):585–594
26. Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S et al (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife. doi:10.7554/eLife.03703
27. Jakočiūnas T, Bonde I, Herrgård M, Harrison SJ, Kristensen M, Pedersen LE et al (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metab Eng 28:213–222
28. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M et al (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res 15:fov004–fov004
29. Stovicek V, Borodina I, Forster J (2015) CRISPR-Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. Metab Eng Commun 2:13–22
30. Storici F, Durham CL, Gordenin DA, Resnick MA (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc Natl Acad Sci 100:14994–14999
31. Kuijpers NGA, Solis-Escalante D, Bosman L, van den Broek M, Pronk JT, Daran J-M et al (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. Microb Cell Fact 12:47
32. Horwitz AA, Walter JM, Schubert MG, Kung SH (2015) Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Syst 1:1–9
33. Donald KA, Hampton RY, Fritz IB (1997) Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. Appl Environ Microbiol 63:3341–3344
34. Yan G-L, Wen K-R, Duan C-Q (2012) Enhancement of β-carotene production by over-expression of HMG-CoA reductase coupled with addition of ergosterol biosynthesis inhibitors in recombinant *Saccharomyces cerevisiae*. Curr Microbiol 64:159–163
35. Verwaal R, Wang J, Meijnen J-P, Visser H, Sandmann G, van den Berg JA et al (2007) High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. Appl Environ Microbiol 73:4342–4350
36. Yamano S, Ishii T, Nakagawa M, Ikenaga H, Misawa N (1994) Metabolic engineering for production of beta-carotene and lycopene in *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 58:1112–1114
37. Janbandhu VC, Moik D, Fässler R (2014) Cre recombinase induces DNA damage and tetraploidy in the absence of loxP sites. Cell Cycle 13:462–470
38. Heider SAE, Peters-Wendisch P, Wendisch VF, Beekwilder J, Brautaset T (2014) Metabolic engineering for the microbial production of carotenoids and related products with a focus on the rare C50 carotenoids. Appl Microbiol Biotechnol 98:4355–4368
39. Vachali P, Bhosale P, Bernstein PS (2012) Microbial carotenoids. Methods Mol Biol 898:41–59
40. Sandmann G (1994) Carotenoid biosynthesis in microorganisms and plants. Eur J Biochem 223:7–24
41. Lee PC, Schmidt-Dannert C (2002) Metabolic engineering towards biotechnological production of carotenoids in microorganisms. Appl Microbiol Biotechnol 60:1–11
42. Ye VM, Bhatia SK (2012) Pathway engineering strategies for production of beneficial carotenoids in microbial hosts. Biotechnol Lett 34:1405–1414
43. Kennedy MA, Bard M (2001) Positive and negative regulation of squalene synthase (ERG9), an ergosterol biosynthetic gene, in *Saccharomyces cerevisiae*. Biochim Biophys Acta 1517:177–189
44. Sherman F (2002) Getting started with yeast. Meth Enzymol 350:3–41
45. Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31–34

Paper 3.

Ronda C., L.E Pedersen, H.G. Hensen, T.B. Kallehauge, M.J. Betenbaugh, A.T. Nielsen, H.F. Kildegaard. **Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool**. *Biotechnol. Bioeng.* **111,** 1604–1616 (2014).

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Accelerating Genome Editing in CHO Cells Using CRISPR Cas9 and CRISPy, a Web-Based Target Finding Tool

Carlotta Ronda,[1] Lasse Ebdrup Pedersen,[1] Henning Gram Hansen,[1] Thomas Beuchert Kallehauge,[1] Michael J. Betenbaugh,[1,2] Alex Toftgaard Nielsen,[1] Helene Faustrup Kildegaard[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark; telephone: +45 2012 4629; fax: +45 4525 8001; e-mail: hef@biosustain.dtu.dk

[2]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland

ABSTRACT: Chinese hamster ovary (CHO) cells are widely used in the biopharmaceutical industry as a host for the production of complex pharmaceutical proteins. Thus genome engineering of CHO cells for improved product quality and yield is of great interest. Here, we demonstrate for the first time the efficacy of the CRISPR Cas9 technology in CHO cells by generating site-specific gene disruptions in COSMC and FUT8, both of which encode proteins involved in glycosylation. The tested single guide RNAs (sgRNAs) created an indel frequency up to 47.3% in COSMC, while an indel frequency up to 99.7% in FUT8 was achieved by applying lectin selection. All eight sgRNAs examined in this study resulted in relatively high indel frequencies, demonstrating that the Cas9 system is a robust and efficient genome-editing methodology in CHO cells. Deep sequencing revealed that 85% of the indels created by Cas9 resulted in frameshift mutations at the target sites, with a strong preference for single base indels. Finally, we have developed a user-friendly bioinformatics tool, named "CRISPy" for rapid identification of sgRNA target sequences in the CHO-K1 genome. The CRISPy tool identified 1,970,449 CRISPR targets divided into 27,553 genes and lists the number of off-target sites in the genome. In conclusion, the proven functionality of Cas9 to edit CHO genomes combined with our CRISPy database have the potential to accelerate genome editing and synthetic biology efforts in CHO cells.

Biotechnol. Bioeng. 2014;111: 1604–1616.

Carlotta Ronda and Lasse Ebdrup Pedersen contributed equally to this work.

The authors declare no conflict of interest.

Correspondence to: H.F. Kildegaard

Contract grant sponsor: The Novo Nordisk Foundation

Received 11 January 2014; Revision received 1 March 2014; Accepted 7 March 2014

Accepted manuscript online 18 May 2014;

Article first published online 22 May 2014 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.25233/abstract).

DOI 10.1002/bit.25233

KEYWORDS: CRISPR Cas9; genome editing; CRISPy; Chinese hamster ovary cells; database

Introduction

Chinese hamster ovary (CHO) cells are the primary factories for biopharmaceuticals, due to their capacity to correctly fold and post-translationally modify recombinant proteins compatible with humans (Jayapal et al., 2007). Genome editing and engineering are of increasing interest in this field for the purpose of increasing cellular production capabilities and improving product quality. This is facilitated by the expanding amount of data being generated for CHO cells including genomic sequences (Lewis et al., 2013; Xu et al., 2011) and other 'omics data such as transcriptomics, proteomics, and metabolomics information (Kildegaard et al., 2013). Early efforts to engineer CHO cells by gene disruptions have been performed mainly by conventional gene targeting strategies based on homologous recombination (HR) (Yamane-Ohnuki et al., 2004). However, HR-based gene targeting is rare event in mammalian cells, since non-homologous end-joining (NHEJ) occurs several orders of magnitude more frequently than HR (Sedivy and Sharp, 1989). NHEJ is an imperfect repair process that often results in insertions or deletions of DNA bases at the site of the double strand break (DSB) during repair, making NHEJ particularly applicable for generating gene disruptions. To induce site specific gene disruptions, targeting endonucleases like transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs), and meganucleases have been

successfully applied to mammalian cells such as human and CHO cell lines (Galetto et al., 2009; Miller et al., 2011; Santiago et al., 2008).

More recently the RNA-guided Cas9 nuclease has proven to be a highly valuable tool for genome editing in nematodes (Waaijers et al., 2013), fruitflies (Bassett et al., 2013; Gratz et al., 2013), zebrafish (Chang et al., 2013; Hwang et al., 2013), plants (Jiang et al., 2013), mice (Wang et al., 2013; Yang et al., 2013), and human cells (Cho et al., 2013; Cong et al., 2013; Mali et al., 2013). Cas9 is the effector protein of the type II clustered regularly interspaced short palindromic repeat (CRISPR) immune system of *Streptococcus pyogenes* and functions as a RNA-guided endonuclease (Carroll, 2012; Jinek et al., 2012). Together with two noncoding RNAs called CRISPR-RNA (crRNA) and trans-activating crRNA (tracrRNA), Cas9 binds to and cleaves DNA in a site-specific manner. The specificity is brought about by the crRNA that basepairs to the target DNA. The target site must be adjacent to a protospacer adjacent motif (PAM) consisting of a random nucleotide and two guanines (NGG) (Jinek et al., 2012; Mali et al., 2013). The tracrRNA molecule together with crRNA functions as a scaffold onto which Cas9 binds. In recent studies, a chimeric RNA that combines the crRNA and tracrRNA termed single guide RNA (sgRNA) has been applied (Chang et al., 2013; Cong et al., 2013; DiCarlo et al., 2013; Jinek et al., 2013; Mali et al., 2013).

In order to express the small chimeric sgRNA, an RNA pol III promoter is required and in previous studies on human cells, a U6 promoter was chosen for this purpose (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). Since the U6 promoter initiates transcription at a guanine (G), this base must be present in the 5′ end of the genomic target site sequence. The U6 promoter-dependent requirement combined with the PAM motif gives rise to the following general scheme of the genomic target site sequence: G $(N)_{19}$NGG. One potential advantage of CRISPR Cas9 technology, when compared to existing methods using TALENs and ZNFs, is its relative low cost. Furthermore, in contrast to ZNFs and TALENs, time-consuming protein engineering is not required to obtain an effective endonuclease (Pennisi, 2013). Thus, the CRISPR technology is attractive for the CHO cell line engineering field and for the genome engineering and synthetic biology community at large.

In this study, we demonstrate for the first time the application of a CHO codon-optimized Cas9 for modifying the genome of CHO cells by disrupting COSMC and FUT8, which are genes encoding proteins involved in O- and N-glycosylation, respectively (Miyoshi et al., 1999; Wang et al., 2010). Deep sequencing analysis revealed a strong preference toward single nucleotide indels. Since no CRISPR bioinformatics tools are available for optimal sgRNA design in CHO cells, the web-based bioinformatics tool "CRISPy" was developed. This design tool facilitates easy and fast sgRNA selection and also incorporates information on possible off-target sites. Combining the CRISPR Cas9 technology and the CRISPy bioinformatics tool, we demonstrate efficient, fast and low cost genetic manipulation of the CHO genome.

Materials and Methods

Plasmid Construction and sgRNA Target Design

The Cas9 sequence from the *S. pyogenes* strain M1 GAS genome with a 3′ nuclear localization signal was codon-optimized for CHO cells, synthesized (for sequence, see Supplementary Materials and Methods) and subcloned into the mammalian expression vector pJ607-03 (DNA 2.0, Menlo Park, CA, Fig. 1A). The plasmid was then transformed into DH5α subcloning cells (Life Technologies, Paisley, U.K.). Transformant clones were selected on 100 μg/mL Ampicillin (Sigma–Aldrich, St. Louis, MO) LB plates. The chosen sgRNA target sequences are listed in Supplementary Table SI. The sgRNA expression constructs were designed by fusing tracrRNA and crRNA into a chimeric sgRNA (Jinek et al., 2012) and located immediately downstream of a U6 promoter (Chang et al., 2013). The sequences of the U6 promoter, scaffold and terminator are shown in Supplementary Materials and Methods. Initially, the sgRNA expression cassette (Fig. 1A) was synthesized as a gBlock (Integrated DNA Technologies, Leuven, Belgium) and subcloned into the pRSFDuet-1 vector (Novagen, Merck, Damstadt, Germany) using KpnI and HindIII restriction sites. This pRSFDuet-1/sgRNA expression vector was used as backbone in a PCR-based uracil-specific excision reagent (USER) cloning method (Hansen et al., 2012; Nour-Eldin et al., 2006). This method was designed to easily and rapidly change the 19 bp-long variable region (N_{19}) of the sgRNA in order to generate our sgRNA constructs. From the pRSFDuet-1/sgRNA expression vector, a 4,221 bp-long amplicon (expression vector backbone) was generated by PCR (1×: 98°C for 2 min; 30×: 98°C for 10 s, 57°C for 30 s, 72°C for 4 min 12 s; 1×: 72°C for 5 min) using two uracil-containing primers (sgRNA Backbone_fw and sgRNA Backbone_rv, Integrated DNA Technologies, Supplementary Table SII) and the X7 DNA polymerase (Nørholm, 2010). Subsequent to Fastdigest DpnI (Thermo Fisher Scientific, Waltham, MA) treatment, the amplicon was purified from a 2% agarose TBE gel using the QIAEX II Gel Extraction Kit (Qiagen, Hilden, Germany). In parallel, 54 bp-long and 53 bp-long single stranded oligos (sense and antisense strand, respectively) comprising the variable region of the sgRNA were synthesized (TAG Copenhagen, Denmark, Supplementary Table SII). The sense and antisense single stranded oligos (100 μM) were annealed in NEBuffer4 (New England Biolabs, Ipswich, MA) by incubating the oligo mix at 95°C for 5 min in a heating block and the oligo mix was subsequently allowed to slowly cool to RT by turning off the heating block. The annealed oligos were then mixed with the gel purified expression vector backbone and treated with USER enzyme (New England Biolabs) according to manufacturer's recommendations. After USER enzyme treatment, the reaction mixture was transformed into *E. coli* Mach1 competent cells (Life Technologies)

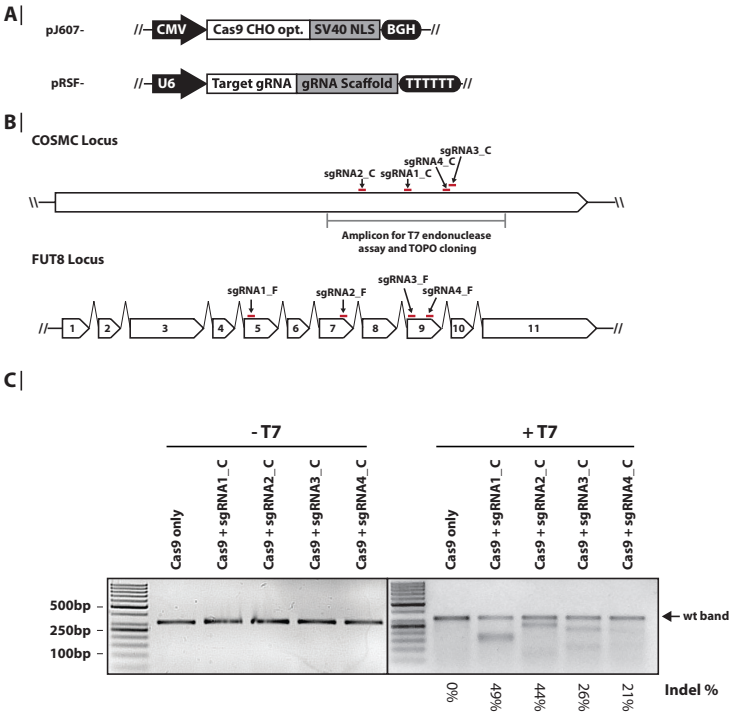


Figure 1. Genome editing in CHO cells by CRISPR Cas9. **A**: Schematics of the Cas9 and sgRNA expression cassettes. The Cas9 expression cassette consists of a CMV promoter, Cas9 ORF codon-optimized for CHO, SV40 nuclear localization sequence (NLS) and bovine growth hormone (BGH) polyadenylation signal and transcription termination sequence. The sgRNA expression cassette consists of a U6 polymerase III promoter, a target gRNA sequence, a gRNA scaffold sequence and a poly(T) termination sequence. **B**: Illustration of the sgRNA genomic target sites in COSMC and FUT8. Red lines denote the position of the sgRNA target sites. Introns are depicted as broken lines (not drawn to scale) and exons as arrowed boxes. **C**: Indel frequency in COSMC analyzed by T7 endonuclease assay. Genomic DNA was extracted from CHO-K1 cells 5 days after transfection with plasmids encoding Cas9 and sgRNA against COSMC. The PCR amplicon covering the sgRNA-target sites as shown in panel B was re-annealed to enable heteroduplex formation before treatment with T7 endonuclease where indicated. Samples were subsequently analyzed by agarose gel electrophoresis. Approximate quantification of indels (%) was peformed with ImageJ software analysis of the uncut (WT) DNA bands. For details see Supplementary Table SVII.

according to standard procedures. Transformant clones were selected on 50 μg/mL Kanamycin (Sigma–Aldrich) LB plates. All constructs were verified by sequencing and purified by NucleoBond Xtra Midi EF (Macherey-Nagel, Düren, Germany) according to manufacturer's guidelines.

Cell Culture and Transfection

CHO-K1 adherent cells obtained from ATCC (#ATCC-CCL-61) were grown in CHO-K1 F-12K medium (ATCC) supplemented with 10% fetal calf serum (Life Technologies) and 1% Penicillin–Streptomycin (Sigma–Aldrich). Cells were expanded in T-75 cm^2 vented cap tissue culture flasks (SARSTEDT, Nümbrecht, Germany) and experiments were performed in Advanced TC Cell Culture Multiwell plates (Greiner Bio-one, Frickenhausen, Germany). Cells were released from plastic ware using trypsin-EDTA (Sigma–Aldrich). Cells were transfected (Day 0) by the Nucleofector 2b device using the Amaxa Cell Line Nucleofector Kit V (Lonza, Basel, Switzerland) according to manufacturer's guidelines (program U-023). A total of 1×10^6 cells were transfected with 1 μg Cas9 plasmid and 1 μg sgRNA plasmid. Cells were incubated at 30°C, 5% CO_2 from Day 1 to Day 2 (cold shock) and incubated at 37°C, 5% CO_2 at all other times. Two days after transfection (Day 2), cells transfected with the pmaxGFP plasmid (Lonza) were used to estimate the

transfection efficiency by analyzing GFP signal using a Celigo Imaging Cell Cytometer (Brooks Automation, Chelmsford, MA). The transfection efficiency was calculated as the percentage of GFP positive cells. Five days after transfection (Day 5), cells were trypsinized and pelleted (200 g, 5 min, RT). Genomic DNA was extracted from the cell pellets using QuickExtract DNA extraction solution (Epicentre, Illumina, Madison, WI) according to manufacturer's instructions and stored at −20°C.

Selection and Phenotypic Analysis of FUT8 Knockout Cells

Five days after transfection (Day 5), selection of FUT8 knockout cells was initiated by supplementing complete medium with 50 μg/mL *Lens culinaris* agglutinin (LCA; Vector Laboratories, Peterborough, UK) from a 5 mg/mL LCA (10 mM Hepes/NaOH, pH 8.5, 0.15 mM NaCl, 0.1 mM $CaCl_2$) stock solution. Bright field images were taken with a Celigo Imaging Cell Cytometer (Brooks Automation). After 7 days of selection (Day 12), genomic DNA was extracted as described above. In parallel, cells were seeded in complete medium without LCA. The day after (Day 13), cells were incubated for 45 min at RT in complete medium containing 20 μg/mL fluorescein-LCA (Vector Laboratories) and two droplets NucBlue® Live ReadyProbes (Life Technologies) per mL media. Cells were washed three times with complete medium and fluorescence microscopy was performed on a LEAP instrument (Intrexon, Germantown, MD) using the two channel imaging application with the NucBlue stain as target 1 using the blue fluorescence channel and the fluorescein labeled LCA as target 2 using the green fluorescence channel.

T7 Endonuclease Assay

Genomic regions flanking the CRISPR target site for T7 endonuclease assay were amplified from the genomic DNA extracts using DreamTaq DNA polymerase (Thermo Fisher Scientific) by touchdown PCR for COSMC (95°C for 2 min; 10×: 95°C for 30 s, 69–59°C (−1°C/cycle) for 30 s, 72°C for 50 s; 20×: 95°C for 30 s, 59°C for 30 s, 72°C for 50 s; 72°C for 5 min), using PCR primers listed in Supplementary Table SII. The PCR products were subjected to a re-annealing process to enable heteroduplex formation which is sensitive to T7 digestion: 95°C for 10 min; 95–85°C ramping at −2°C/s; 85°C to 25°C at −0.25°C/s; and 25°C hold for 1 min. Re-annealed PCR products were treated with T7 endonuclease (New England Biolabs) for 30 min at 37°C. T7 digested and undigested samples were analyzed on a 3% TAE gel. The percentage of indels was estimated from analysis of the uncut (WT) gel bands with ImageJ software. For details see Supplementary Table SVII.

TOPO™ TA Cloning and Sanger Sequencing

A genomic region of 318 bp covering the four COSMC sgRNA target sites was PCR-amplified from the genomic extracts as described in the T7 endonuclease assay. PCR products were subjected to agarose gel electrophoresis and subsequently gel purified from a 1% agarose TBE gel using the QiaQuick Gel Extraction Kit (Qiagen). Purified PCR products were TOPO-cloned into the pCR4-TOPO vector using the TOPO™ TA cloning kit (Life Technologies) and subsequently transformed into *E. coli* Mach1 chemically competent cells (Life Technologies). Transformed Mach1 cells were then plated on LB-ampicillin agar plates and grown at 37°C overnight. Plasmids from single colony 60 μg/mL carbenicillin (Novagen, Merck) 2X YT-cultures were extracted using the Nucleospin 8/96 Plasmid kit (Macherey-Nagel). Each plasmid preparation was sequenced using the M13 forward (−20) primer (Supplementary Table SII) on an AB 3500xL Genetic Analyzer (Life Technologies) using the BigDye Terminator v3.1 cycle sequencing kit (Life Technologies).

MiSeq Library Construction and Deep Sequencing

PCR amplicons were designed to be between 150 and 200 bp long and to span the sgRNA target sequence (See Supplementary Table SII for primers and Supplementary Table SIII for amplicon sizes). Amplicons were generated from the genomic DNA extracts using Phusion Hot Start II HF Pfu polymerase (Thermo Fisher Scientific) by touchdown PCR (95°C for 7 min; 20×: 95°C for 45 s, 69°C − 59°C (−0.5°C/cycle) for 30 s, 72°C for 30 s; 35×: 95°C for 45 s, 59°C for 30 s, 72°C for 30 s; 72°C for 7 min). Amplicons were purified on 2% agarose TBE gels and bands with expected fragment sizes were excised and purified using QIAEX II Gel Extraction Kit (Qiagen). Amplicon concentration was measured on Qubit® using the dsDNA BR Assay Kit (Life Technologies). Amplicons were pooled in four for multiplexing (25 ng each, 100 ng in total). Illumina multiplexing adapters were ligated to the pooled amplicons using the TruSeq™ LT DNA Sample Preparation LT kit (Illumina) according to manufacturer's instructions. DNA concentration of the multiplexed libraries was measured with the Qubit® dsDNA BR Assay Kit, and library quality was determined with an Agilent DNA1000 Chip (Agilent Bioanalyzer 2100). Finally, multiplexed libraries were pooled and sequenced on a MiSeq Benchtop Sequencer (Illumina) using the MiSeq Reagent Kit v2 (300 cycles) according to manufacturer's protocol for a 151 bp paired-end analysis.

Deep Sequencing Data Analysis

To minimize the number of required indexes, the same index was used on different PCR products (multiplexing) and the identities of the PCR products were found in the data analysis step based on their individual PCR primer sequences. A Python script was developed to process MiSeq data resulting from the targeted re-sequencing of the Cas9 target site regions. The script performs the following tasks: (1) join paired-end reads; (2) check if resulting sequences contain correct PCR primer at both beginning and end of the

sequences and discard those sequences that fail to do so; (3) compute output length of PCR product; and (4) compare PCR product length to expected PCR product length. Paired-end reads were joined using fastq-join (Aronesty 2011: http://code.google.com/p/ea-utils). Ends were checked for correct PCR primer using fastx_barcode_splitter (http://hannonlab.cshl.edu/fastx_toolkit/index.html).

Cas9 Target Finding Database and Web Interface CRISPy

A python script, utilizing BioPython (Cock et al., 2009), was created to search through the CHO-K1 genome obtained from http://www.chogenome.org (Hammond et al., 2012). Potential target sequences of the format $G(N)_{19}NGG$ were searched for in annotated exons. Each of the identified target sequences was then searched against the entire genome for potential off-targets. Only genomic sequences matching the 13 bp sequence immediately upstream of the NGG were identified as potential off-targets, and one or two mismatches were allowed. Additionally, it was tested if an off-target/mismatch was located within an exon. The database generated by the Python script was uploaded to a MySQL database and an interface based on HTML, PHP, and JavaScript was created to allow public access to the database.

Results

RNA-Guided CRISPR Cas9 Shows Targeted Endonuclease Activity in CHO

In order to test if the RNA-guided CRISPR Cas9 system could be applied for gene disruptions in CHO cells, an expression vector with a CHO codon-optimized version of Cas9 with a C-terminal SV40 nuclear localization signal under the control of a CMV promoter was constructed (Fig. 1A). To direct Cas9 to disrupt genes of interest, sgRNA expression constructs were generated using the human U6 polymerase III promoter as previously described (Mali et al., 2013) (Fig. 1A). Four sgRNAs were designed for each of the two genes; C1GALT1C1 (COSMC) encoding the C1GALT1-specific chaperone 1 and FUT8 encoding fucosyltransferase 8 (alpha-(1,6)-fucosyltransferase). COSMC is a chaperone essential for correct protein O-glycosylation (Wang et al., 2010) and FUT8 catalyzes the transfer of fucose from GDP-fucose to N-acetylglucosamine (Wilson et al., 1976). In general, it may be desirable to choose a target early in the gene in order to avoid a truncated yet partially functional protein. However, knowledge regarding alternative splicing or active sites may in some cases make a more downstream position or exon a better choice. The four sgRNA constructs for COSMC target the only exon present in the gene. This exon has previously been targeted with a ZFN in human cells (Steentoft et al., 2011). FUT8 consists of 11 exons and the FUT8 sgRNA constructs target exon 5, exon 7, and exon 9 (Fig. 1B). Exon 9 was chosen based on previously published work targeting FUT8 with a ZFN (Malphettes et al., 2010), and exons 5 and 7 were chosen to target earlier exons in the gene sequence. To compare the activity of the designed sgRNAs, adherent CHO-K1 cells were transfected transiently with the CHO codon-optimized Cas9 expression vector and each of the eight sgRNAs to introduce DSBs in the two test genes in two independent experiments (replicate 1 and 2). Initially, a T7 endonuclease assay was performed to analyze the indel frequency at the COSMC loci resulting from Cas9 guided by the four different COSMC-targeting sgRNAs (sgRNA1_C, sgRNA2_C, sgRNA3_C, and sgRNA4_C). When assayed 5 days after transfection, genomic indel events were detected for all four sgRNAs (Fig. 1C, replicate #1 is shown). Using ImageJ software analysis of the uncut gel bands, the percentage of indels generated at the COSMC loci was estimated to be between 21% and 49% (Supplementary Table SVII). The fragment sizes of the digested amplicons correspond to the expected sizes (Supplementary Table SIV).

High Indel Frequency Obtained by All Four COSMC-Targeting sgRNAs

To further assess the indel frequency achieved with the COSMC sgRNAs, TOPO cloning-based sequencing of gel-purified amplicons from the COSMC genomic site was performed (Fig. 2A, Supplementary Table SV). Consistent with the T7 endonuclease assay, Cas9 activity was observed for all four COSMC sgRNAs in two independent experiments (replicate 1 and 2). sgRNA1_C gave rise to the highest indel frequency of 48.0% and 66.7% and sgRNA4_C displayed the lowest indel frequency of 10.3% and 17.9%, in replicate 1 and 2, respectively. Based on GFP fluorescence of cells transfected with GFP-encoding plasmids, transfection efficiency was estimated to be approximately 60% and 65% for replicate 1 and 2, respectively. The indels created by the COSMC sgRNAs predominantly involved a single-base insertion of a thymine or deletions (Fig. 2B; only sgRNA1_C, replicate #1 is shown). To analyze the Cas9 activity in greater detail, deep sequencing was performed using the genomic DNA extracts from the two independent experiments (Fig. 2C). Deep sequencing data comprising between approximately 200,000–700,000 reads per sgRNA in each of the two replicates correlated well with the sequencing data obtained from TOPO cloning (between 21 and 32 sequences per sgRNA). Both sequence-based methods detected relatively high Cas9-activity for all four sgRNAs. Deep sequencing reported indel frequencies of 47.3% and 44.3% for sgRNA1_C, 45.6% and 40.2% for sgRNA2_C, 36.0% and 27.2% for sgRNA3_C and 15.2% and 13.6% for sgRNA4_C in replicate 1 and 2, respectively. Deep sequencing of control cells transfected only with Cas9-encoding plasmids showed an indel frequency of 0.1–0.2% (Supplementary Fig. S1). To examine the fidelity of both sequence-based methods, indel-containing sequences obtained from TOPO-cloning were checked using the deep sequencing data. All indels detected in the TOPO-cloning experiments were also retrieved in the deep sequencing data (data not shown).

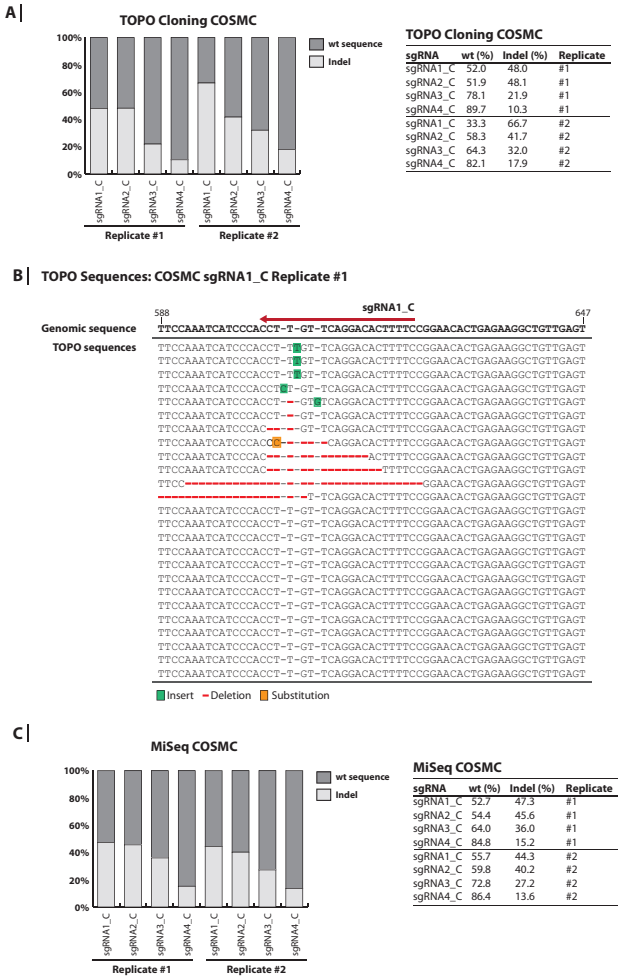


Figure 2. Analysis of generated indels in COSMC. **A**: TOPO™ TA-based sequence analysis of COSMC. Genomic DNA was extracted from CHO-K1 cells 5 days after transfection with plasmids encoding Cas9 and sgRNA against COSMC. PCR amplicons covering the sgRNA-target sites in COSMC were TOPO™ TA-cloned and Sanger sequenced. Between 21 and 32 sequences were obtained for each sgRNA. The percentages of wt and indel sequences are illustrated in the bar plot and shown in the table. **B**: Alignment of TOPO™ sequence traces. Genomic DNA from cells transfected with Cas9 + sgRNA1_C (replicate #1) were subjected to TOPO™ cloning as described in panel A. The sequence traces are aligned to the CHO-K1 genomic sequence. The red arrow indicates the genomic target site of sgRNA1_C. The numbers denote the position (bp) in the open reading frame of COSMC. Green, red and orange colors indicate insertions, deletions and substitutions, respectively. **C**: Targeted deep sequencing analysis of COSMC. The same extracted genomic DNA as described for panel A was used as template for the MiSeq analysis. Between 200,000 and 700,000 sequences were obtained for each sgRNA. The percentages of wt and indel sequences are illustrated in the bar plot and shown in the table.

Homozygous Knockout of FUT8 in CHO Cells Generated by CRISPR

The α1,6-fucosyltransferase FUT8 catalyzes the addition of fucose on IgG1 antibodies produced by CHO cells which can reduce antibody-dependent cell-mediated cytotoxicity (Niwa et al., 2004; Shields et al., 2002; Shinkawa et al., 2003). Disruption of the FUT8 gene in CHO cells is therefore attractive in order to achieve highly active and completely nonfucosylated therapeutic antibodies (Yamane-Ohnuki et al., 2004). To expand our knowledge of applying CRISPR Cas9 in CHO cells, the gene disruption efficiency of four FUT8 sgRNAs was investigated by deep sequencing. Genomic regions covering the target site of sgRNA1_F, sgRNA2_F, sgRNA3_F, and sgRNA4_F were PCR amplified and sequenced. This analysis revealed that all four sgRNAs gave rise to significant Cas9 activity with an indel frequency of 17.6% and 15.1% for sgRNA1_F, 38.7% and 31.2% for sgRNA2_F, 42.5% and 36.0% for sgRNA3_F, and 18.9% and 11.1% for sgRNA4_F in replicate 1 and 2, respectively (Fig. 3A). As previously mentioned, transfection efficiency was estimated to be approximately 60% and 65% for replicate 1 and 2, respectively. *Lens culinaris* agglutinin (LCA)-based selection was further used to select for FUT8-disrupted CHO cells. LCA binds fucosylated plasma membrane proteins leading to endocytosis and cell death. This enables selection for homozygous FUT8 gene disruptions, since LCA can no longer bind to cells devoid of FUT8 and these cells therefore survive (Malphettes et al., 2010). LCA-treatment was initiated 5 days after transfection and resulted in non-adherent round-shaped morphology of all control cells (Fig. 3B). However, many adherent cells were detected in the pool of cells transfected with Cas9 and the four FUT8 sgRNAs (Fig. 3B, only sgRNA3_F is shown), indicating Cas9-mediated functional knockout of FUT8 in these cells. To analyze the phenotypic change of CRISPR Cas9 mediated disruption of FUT8 on cell surface exposed α-1,6-linked fucose moieties, a lectin stain was performed (Malphettes et al., 2010; Mori et al., 2004; Yamane-Ohnuki et al., 2004). Eight days after initiation of selection, LCA selected and non-LCA selected cells transfected with and without Cas9 and sgRNAs were stained with fluorescein-labeled LCA (F-LCA) (Fig. 3C and Supplementary Fig. S2). Cells transfected with Cas9 + sgRNAs without LCA selection revealed a fraction of F-LCA negative cells, demonstrating the presence of cells with homozygous disruption of the FUT8 gene. For Cas9 + sgRNA3_F, these F-LCA negative cells constituted 29.1% of the entire population (Fig. 3D). Cells transfected with Cas9 + sgRNA3_F, which subsequently had been exposed to LCA treatment revealed that the majority of cells (98.6%) stained LCA negative (Fig. 3D). This clearly demonstrates that the LCA treatment efficiently selects for cells devoid of functional FUT8 as previously observed (Malphettes et al., 2010, Mori et al., 2004; Yamane-Ohnuki et al., 2004). Indeed, this observation was confirmed by deep sequencing, since LCA selection significantly enriched cells with FUT8 disruption to an indel frequency between 98.2% and 99.7% for the four FUT8-targeting sgRNAs (Fig. 3A).

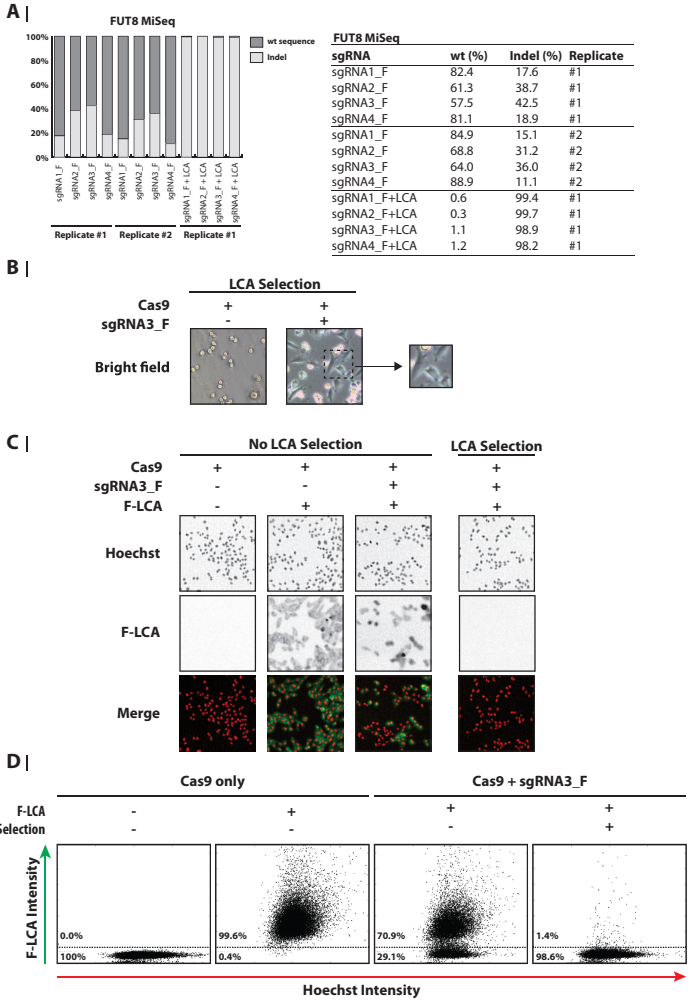
The Majority of CRISPR-Generated Indels Is Single Base Pair Insertions

The vast amount of information obtained from deep sequencing led us to investigate further the indel sizes created by the NHEJ repair mechanism resulting from the COSMC and FUT8 sgRNAs. The frequency was calculated as an average for both independent experiments and was based on 11×10^5 reads for COSMC sgRNAs and 8×10^5 reads for FUT8 sgRNAs (Supplementary Figs. S3 and S4 and Table SVI). All targets were weighted equally with each target contributing 12.5%. The data was compiled into a single plot, displaying the frequency of specific indel sizes ranging from 37 bp deletions to 11 bp insertions within the individual targets (Fig. 4A). Surprisingly, mainly single base pair insertions were detected with a frequency of 32.8%. A high frequency of single base pair insertions was also observed in sequences obtained from TOPO cloning (Fig. 2B, only sgRNA1_C, replicate #1 is shown). Two and one base pair deletions were the second and third most frequent indel size with a frequency of 10.3% and 8.7%, respectively. Together, almost half of the identified indels (56.5%) were single or double-base pair indels. Collectively, 85% of the indels observed in this study resulted in frame shift mutations (± 1 or ± 2 bp) in the reading frame (Fig. 4B), which most likely leads to a loss-of-function of the target protein. This finding further underlines CRISPR Cas9 as a powerful tool to disrupt genes of interest in the CHO genome, and prompted us to develop a target design tool that facilitates identification of Cas9 targets.

Cas9 Target Finding Tool CRISPy for CHO-K1

To our knowledge, there are currently three Cas9 target design tools available to the public. The "ZiFiT Targeter" (http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx) identifies Cas9 targets in a given sequence. The "CRISPR Design" (http://crispr.mit.edu/) (Hsu et al., 2013), allows the user to find targets in a given sequence and then checks for off-targets in the genome of either human, mouse, zebrafish, or *C. elegans*. The "Cas9 guide RNA Design" tool (Ma et al., 2013) is highly similar to "CRISPR Design" with the addition of reporting content of AT (adenine and thymine) and predicting secondary RNA structure. However, neither can currently be applied to find sgRNA target sequences with off-target information in CHO genomes, nor provide pre-configured links to primer design tools.

In order to design the Cas9 system for gene knockouts, a number of tasks must be performed. Once a gene of interest has been selected, one or more targets must be identified in the exons of the gene. These targets should then be evaluated based on their position in the gene and sequence similarity to the rest of the genome. Once suitable targets have been



FUT8 MiSeq

sgRNA	wt (%)	Indel (%)	Replicate
sgRNA1_F	82.4	17.6	#1
sgRNA2_F	61.3	38.7	#1
sgRNA3_F	57.5	42.5	#1
sgRNA4_F	81.1	18.9	#1
sgRNA1_F	84.9	15.1	#2
sgRNA2_F	68.8	31.2	#2
sgRNA3_F	64.0	36.0	#2
sgRNA4_F	88.9	11.1	#2
sgRNA1_F+LCA	0.6	99.4	#1
sgRNA2_F+LCA	0.3	99.7	#1
sgRNA3_F+LCA	1.1	98.9	#1
sgRNA4_F+LCA	1.2	98.2	#1

Figure 3. Functional and genomic knockout of FUT8 in CHO-K1. **A**: Targeted deep sequencing analysis of the FUT8 locus in CHO-K1 cells. Genomic DNA was extracted from CHO-K1 cells transfected with Cas9 and FUT8 sgRNAs harvested on Day 5 (no LCA selection) or Day 12 (7 days with LCA selection). The percentages of wt and indel sequences are illustrated in the bar plot and shown in the table. **B**: Selection of FUT8 knockout CHO-K1 cells by LCA. As indicated, cells were either transfected with only a Cas9-encoding plasmid or in combination with an sgRNA3_F-encoding plasmid. Five days after transfection (Day 5), selection with LCA was initiated. The day after (Day 6), the shown bright field images were acquired. The magnified view shows cells with normal (adherent-looking) morphology from pool of cells transfected with Cas9 and sgRNA3_F. **C**: Phenotypic staining of FUT8 knockout CHO-K1 cells by fluorescein-labeled LCA (F-LCA). CHO-K1 cells were treated as described for panel A. On Day 13, cells were treated with Hoechst and with F-LCA where indicated. Fluorescence microscopy images were subsequently acquired. Hoechst and F-LCA signal is depicted as red and green color, respectively, in the merged images and as grayscale in the individual images. **D**: Quantification of fluorescent-based phenotypic staining of FUT8 knockout CHO-K1 cells. Cells were gated based on signal intensity of Hoechst and F-LCA as shown. The percentages of F-LCA positive (FUT8 WT) and negative (phenotypic knockout of FUT8) cells are shown. LCA: *Lens culinaris* agglutinin.

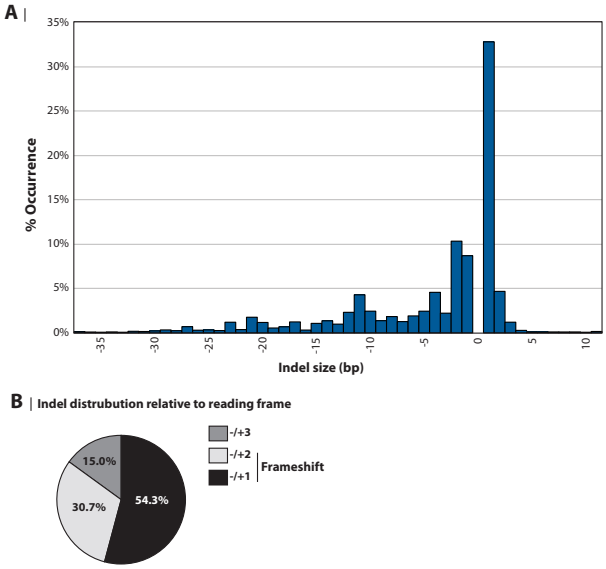


Figure 4. Cas9 activity in CHO results in high frequency of 1 bp indels. **A**: Frequency distribution of indel sizes. The size distribution of indels is based on sequences obtained from all eight sgRNAs. The indel frequency for each sgRNA is the average for both independent experiments. Normalization was performed so each of the eight sgRNA accounts for 12.5% of the data points. Only indels ranging from −37 bp (37 bp deletion) to +11 bp (11 bp insertion) are shown. **B**: Frameshifts generated by CRISPR Cas9. The distribution of indels generating ±1, ±2, and ±3 base pair shifts in the reading frame was calculated from data presented in panel A.

identified, a method for monitoring creation of gene disruptions must be established. This is commonly done using PCR amplification and sequencing of the targeted genomic region. To facilitate this workflow (Fig. 5), we have developed a bioinformatics tool "CRISPy" that is freely accessible at "http://staff.biosustain.dtu.dk/laeb/crispy/." The web interface interacts with a precompiled database of all possible Cas9 target sites in CHO-K1 genes based on the annotated CHO-K1 genome (Hammond et al., 2012; Xu et al., 2011). Every target sequence has the following format $GN_{19}NGG$. For every Cas9 target sequence, we have compiled a list of off-targets of the format $B_{13}NGG$ (B is a nucleotide identical to the genomic sequence and N is a random nucleotide) with 0–2 mismatches. This list also shows the genomic site of each off-target for the 0 and 1 bp mismatches. Overall, this has resulted in a database with 1,970,449 targets divided into 27,553 genes.

While improvements of the CHO-K1 genome annotation are still underway, there are 21,610 coding sequences annotated as of this publication. They contain 231,866 exons. 221,353 of these (95.5%) have at least one potential Cas9 target site. Given CRISPy's 1,970,449 potential Cas9 target sites in exons, this gives an average of nine target sites per exon. If a researcher chooses one of these target sites at random, there is a risk of choosing a target site with high sequence similarity to other parts of the genome; thus producing potential off-target effects. The median number of exact DNA sequence matches elsewhere in the genome for a target site is 6. For 1 bp mismatches the number is 377 and for 2 bp mismatches the number is 4558. Off-targets are here based on the first 13 bp upstream of the PAM sequence. However, 248,777 of 1,970,449 target sites (12.6% unique sites) have zero exact off-target matches. This highlights the importance of using CRISPy to select the most specific CRISPR target site.

The user can search for a gene of interest based on the annotation available for the CHO-K1 genome at the time of database generation; for example, GeneID, gene symbol and name (Fig. 5). After clicking the gene of interest, the user will be presented with a schematic overview of the gene and targets. All exonic targets can either be listed together or on an exon-by-exon basis. Targets are listed with number of off-targets with 0, 1, or 2 bp mismatches as well as a list of genes in which one or more of the perfectly matching off-

Figure 5. Workflow of Cas9 target finding tool: CRISPy. Screenshots of the online CRISPy tool showing the process of (1) finding a target gene, (2) selecting an exon, (3) evaluating the available targets, and (4) showing links to PCR primer designs for analysis of on- and off-target effects resulting from Cas9 + sgRNA activity. The tool is available at http://staff.biosustain.dtu.dk/laeb/crispy/.

targets occur. Once one or more suitable targets have been selected, the user can click on a link to get to a target specific page with direct pre-configured links to the NCBI Primer-Blast tool (Ye et al., 2012) on which the user simply has to click the "Get Primers" button. Once the Primer-Blast tool returns results, the user can verify that the PCRs amplify the desired region by copy/pasting the target sequence into the "Find on Sequence" field, which should return at least one sequence hit (if more than one hit, then select the one at the position indicated as your PCR region). It should be noted

that the pre-configured link to the Primer-Blast tool is set to generate PCR amplicons of 100–200 bp for compatibility with deep sequencing such as MiSeq. The user can alter the size on the Primer-Blast tool page if necessary. In addition there are links to the genomic sequence at either NCBI or UDEL. On the target specific page the user will also find a list of 0 or 1 bp mismatch off-targets for which there are also links to genome sequence and Primer-Blasts to facilitate easy monitoring of potential off-targets. Since the database is precompiled, there are no time-consuming computational steps.

Discussion

In this study, we demonstrate the successful application of RNA-guided CRISPR Cas9 for generating gene disruptions in CHO-K1 cells. The tested sgRNAs for COSMC created indels with a frequency between 13.6% and 47.3% according to MiSeq analysis in a pool of transfected cells with a transfection efficiency of approximately 60%. In comparison, genetic disruption frequency of 3.8% and 6.3% in CHO cells for BAK and BAX, respectively, has been observed using pre-screened ZFNs (Cost et al., 2010). With an indel frequency between 11.1% and 42.5% created at the target sites, the tested sgRNAs for FUT8 revealed an activity similar to the high efficiencies observed for COSMC. Selection pressure with LCA furthermore facilitated enrichment of cells exhibiting functional disruptions in the FUT8 gene for each of the four sgRNAs. Together, RNA-guided Cas9 activity was able to generate indels with a relatively high frequency for all eight sgRNAs examined, demonstrating that the Cas9 genome-editing methodology is robust and efficient. With these high efficiencies obtained with CRISPR Cas9 system in CHO cells, it will be worthwhile to investigate the capacity of Cas9-based multiplexing to generate multiple gene disruptions in a single round of modifications. Since Cas9-based multiplexing has successfully been performed in other mammalian cells (Cong et al., 2013; Wang et al., 2013), multiplexing using the Cas9 system may as well be a powerful technique in CHO cells.

The mutations created by the eight sgRNAs were predominantly very short indels (56.5% single or double base pair indels) with a preference for single base pair insertions. Analysis of indel sizes obtained with Cas9 in human cells revealed mainly single base pair deletions (Mali et al., 2013). The preference for small single base pair deletions was also observed in another study involving Cas9 in human cells (Wang et al., 2014). Interestingly, the preference for single base pair insertions or deletions observed in our study resulted in a high frequency of indels creating frameshifts (85%) within the open reading frame further supporting Cas9 as a highly attractive endonuclease for generating gene disruptions.

To enable high throughput automated gene disruptions in CHO, the bioinformatics tool "CRISPy" was developed to assist in identification of sgRNA target sites. The sgRNA design tool incorporates additional elements/properties not currently available elsewhere including visualization of sgRNA target sites, detailed off-target information and links to primer design tools. Since off-target indel events have been observed in previous reports on Cas9 in human cells (Fu et al., 2013; Hsu et al., 2013; Wang et al., 2014), prescreening sgRNAs for possible off-target effects represents a useful addition to the target design tool box. The CRISPy tool presented here provides upfront off-target analysis of the designed sgRNAs, enabling selection of sgRNAs with the minimal number of possible off-target sites. Furthermore, CRISPy aids researchers in primer design for targeted analysis of off-target effects. We envision incorporating new knowledge on target sequence-dependent activity of sgRNAs as it becomes available in the future.

In this study, we have been able to enrich the FUT8 knockout population by LCA selection. However, this type of selection is often unavailable and so single cell cloning will be required to obtain cells with the desired gene disruptions. This is commonly achieved through either FACS sorting or limited dilution. These clones must then be analyzed for homozygous populations through screening by fragment analysis and sequencing. However, this process can be time consuming and includes a number of challenges. With the high genome editing efficiency of the Cas9 system, the number of analyzed single cell clones sufficient for obtaining homozygous mutations is expected to be lowered considerably. Thus, the CRISPR Cas9 system holds the potential to significantly decrease the heavy workload involved in generating knockout CHO cell lines.

Our study demonstrates that design and implementation of Cas9-sgRNA-based genome engineering is straightforward and fast. Additionally, the CRISPR Cas9 system is relatively inexpensive as the Cas9 expression vector is reused and only new sgRNA constructs need to be cloned for every target sequence at the cost of a few oligonucleotides. The high efficiency, robustness, ease of use, and low costs make the CRISPR Cas9 system a highly attractive genome-editing tool for both the academic and industrial community. The introduction of the CRISPR Cas9 system in CHO cells combined with the CRISPy design tool will significantly accelerate the pace of genome editing in CHO cells and enhance the rate of CHO cell line improvement for increasing yields and quality of biopharmaceuticals.

The authors thank Jae Seong Lee and Bjørn Voldborg for valuable guidance, support, and fruitful discussions. The authors thank Patrice Menard and Sara Petersen Bjørn for technical assistance. In addition, the authors thank Anna Koza for her assistance with the MiSeq experiments. This work was supported by The Novo Nordisk Foundation.

References

Bassett AR, Tibbit C, Ponting CP, Liu J-L. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4:220–228.

Carroll D. 2012. A CRISPR approach to gene targeting. Mol Ther 20:1658–1660.

Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, Xiong J-W, Xi JJ. 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 23:465–472.

Cho SW, Kim S, Kim JM, Kim J-S. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823.

Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, Miller JC, Rebar E, Collingwood TN, Snowden A, Gregory PD. 2010. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng 105:330–340.

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826.

Galetto R, Duchateau P, Pâques F. 2009. Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin Biol Ther 9:1289–1303.

Gratz SJ, Cummings, AM, Nguyen JN, Hamm DC, Donohue LK, Harrison JW, O'Connor-Giles KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194:1029–1035.

Hammond S, Kaplarevic M, Borth N, Betenbaugh MJ, Lee KH. 2012. Chinese hamster genome database: An online resource for the CHO community at w w w.chogenome.org. Biotechnol Bioeng 109:1353–1356.

Hansen BG, Sun XE, Genee HJ, Kaas, CS, Nielsen, JB, Mortensen UH, Frisvad, JC, Hedstrom, L. 2012. Adaptive evolution of drug targets in producer and non-producer organisms. Biochem J 441:219–226.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31:827–832.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh J-RJ, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229.

Jayapal KP, Wlaschin KF, Hu W-S, Yap MGS. 2007. Recombinant protein therapeutics from CHO cells—20 years and counting. Chem Eng Prog 103:40–47.

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41:1–12.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821.

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. eLife 2:1–9.

Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. 2013. The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. Curr Opin Biotechnol 24:1–6.

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili I, Xu X, Wang J, Palsson BO. 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. Nat Biotechnol 31:759–765.

Ma M, Ye AY, Zheng W, Kong L. 2013. A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. BioMed Res Int 2013:270805.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. Science 339:823–826.

Malphettes L, Freyvert Y, Chang J, Liu P-Q, Chan E, Miller JC, Zhou Z, Nguyen T, Tsai C, Snowden AW, Collingwood TN, Gregory PD, Cost GJ. 2010. Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. Biotechnol Bioeng 106:774–783.

Miller JC, Tan S, Qiao G, Barlow K, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ. 2011. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29:143–148.

Miyoshi E, Noda K, Yamaguchi Y, Inoue S, Ikeda Y, Wang W, Ko JH, Uozumi N, Li W, Taniguchi N. 1999. The alpha1-6-fucosyltransferase gene and its biological significance. Biochem Biophys Acta 1473(1):9–20.

Mori K, Kuni-Kamochi R, Yamane-Ohnuki N, Wakitani M, Yamano K, Imai H, Kanda Y, Niwa R, Iida S, Uchida K, Shitara K, Satoh M. 2004. Engineering Chinese hamster ovary cells to maximize effector function of produced antibodies using FUT8 siRNA. Biotechnol Bioeng 88:901–908.

Niwa R, Shoji-hosaka E, Sakurada M. 2004. Defucosylated chimeric anti-CC chemokine receptor 4 IgG1 with enhanced antibody-dependent cellular cytotoxicity shows potent therapeutic activity to T-cell leukemia and lymphoma. Cancer Res 64:2127–2133.

Nour-Eldin, HH, Hansen BG, Nørholm, MHH, Jensen, JK, Halkier, BA. 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res 34:e122.

Nørholm, M. 2010. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol 10:21.

Pennisi E. 2013. The CRISPR craze. Science 341:833–836.

Santiago Y, Chan E, Liu P-Q, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci USA 105:5809–5814.

Sedivy JM, Sharp PA. 1989. Positive genetic selection for gene disruption in mammalian cells by homologous recombination. Proc Natl Acad Sci USA 86:227–231.

Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SH, Presta LG. 2002. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. J Biol Chem 277:26733–26740.

Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N, Shitara K. 2003. The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. J Biol Chem 278:3466–3473.

Steentoft, C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB, Clausen H. 2011. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. Nat Methods 8:977–982.

Waaijers S, Portegijs V, Kerver J, Lemmens BB, Tijsterman G, van den Heuvel S, Boxem M. 2013. CRISPR/Cas9-targeted mutagenesis in *C. elegans*. Genetics 195:1187–1191.

Wang Y, Ju T, Ding X, Xia B, Wang W, Xia L, He M. 2010. Cosmc is an essential chaperone for correct protein O-glycosylation. Proc Natl Acad Sci USA 107:9228–9233.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343:80–84.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153:910–918.

Wilson JR, Williams D, Schachter H. 1976. The control of glycoprotein synthesis: N-acetylglucosamine linkage to a mannose residue as a signal for the attachment of L-fucose to the asparagine-linked N-acetylglucosamine residue of glycopeptide from alpha1-acid glycoprotein. Biochem Biophys Res Commun 72:909–916.

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Chen W, Xie M, Wang W, Hammond S, Mikael R, Neff N, Passarelli B, Koh W, Fan HC, Gui Y, Lee

KH, Betenbaugh MJ, Quake SR. 2011. The genomic sequence of the Chinese hamster ovary (CHO) K1 cell line. Nat Biotechnol 29:735–741.
Yamane-Ohnuki N, Kinoshita S, Inoue-Urakubo M, Kusunoki M, Iida S, Nakano R, Wakitani M, Niwa R, Sakurada M, Uchida K, Shitara K, Satoh M. 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. Biotechnol Bioeng 87:614–622.
Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154:1370–1379.
Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Patent

Ronda C., Pedersen L.E., Kildegaard H.F., Lee J.S. **"Multiplex editing system**". EPO Patent Application (WO2015052231).

Extract from the Register of European Patents

About this file: WO2015052231

WO2015052231 - MULTIPLEX EDITING SYSTEM [Right-click to bookmark this link]			
Status	The international publication has been made *Database last updated on 26.08.2015*		
Most recent event	03.07.2015	PCT data prior to European publication	
Applicant(s)	For all designated states Technical University of Denmark Anker Engelundsvej 1 Bygning 101A 2800 Kgs. Lyngby / DK		
	[N/P]		
Inventor(s)	01 / PEDERSEN, Lasse Ebdrup Mjolnerparken 54 2.th. DK-2200 Kobenhavn N / DK		
	02 / RONDA, Carlotta Sankt Pauls Gade 3 1.tv. DK-1313 Kobenhavn K / DK		
	03 / KILDEGAARD, Helene Faustrup Flintholm Allé 24 3.tv. DK-2000 Frederiksberg / DK		
	04 / LEE, Jae Seong Christianslund 39 2.mf. DK-3460 Birkerød / DK		
	[N/P]		
Application number, filing date	14798710.1	08.10.2014	
	WO2014EP71534		
Priority number, date	EP20130187681	08.10.2013 Original published format: EP 13187681	
	EP20140157146	28.02.2014 Original published format: EP 14157146	
	EP20140169256	21.05.2014 Original published format: EP 14169256	
Filing language	EN		
Procedural language	EN		
Publication	**Type:**	A2 Application without search report	
	No.:	WO2015052231	
	Date:	16.04.2015	
	Language:	EN	
	The application has been published by WIPO in one of the EPO official languages on 16.04.2015		
	[2015/15]		
International and Supplementary search report(s)	**International search report - published on:**	EP	02.07.2015
Classification	**International:**	C12N15/90, C12N15/10, C12N9/22	
Designated contracting states	AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LI, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR		
Extension states	BA	Not yet paid	
	ME	Not yet paid	
Title	**German:**		
	English:	MULTIPLEX EDITING SYSTEM	
	French:	SYSTÈME D'ÉDITION MULTIPLEX	
Cited in	**International search**	XP055067747 A 00000000 [ ] **DOI:** http://dx.doi.org/10.1126/science.1225829	
		XP055067744 A 00000000 [ ] **DOI:** http://dx.doi.org/10.1126/science.1231143	
		[X] 1-2,13-15,19,21-24,26 the whole document, in particular Fig. 1	
		[X] 1-2,19,21-26 the whole document, in particular Fig. 1 and Table 1	
		[X] 19-22 the whole document, in particular Fig. 1 and 2	
		[X] 19-22 p. W	
		[X] 19-22 p. W	
		[X] 19-22 p. W	
		[PX] 1-26 p. W	